

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
15 February 2001 (15.02.2001)

PCT

(10) International Publication Number  
**WO 01/11010 A2**

- (51) International Patent Classification<sup>7</sup>: C12N Claiborne [US/US]; 46 Crabapple Lane, Guilford, CT 06437 (US).
- (21) International Application Number: PCT/US00/21097
- (22) International Filing Date: 2 August 2000 (02.08.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/147,860 9 August 1999 (09.08.1999) US
- (71) Applicant (for all designated States except US): GENAIS-  
SANCE PHARMACEUTICALS, INC. [US/US]; Five  
Science Park, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): CHEW, Anne  
[US/US]; 837 Orange Street, Apartment 2, New Haven,  
CT 06511 (US). DENTON, R., Rex [US/US]; 129 Hunters  
Trail, Madison, CT 06443 (US). DUDA, Amy [US/US];  
93 Marvel Road, New Haven, CT 06515 (US). KLIEM,  
Stefanie, E. [DE/US]; 1298 Hartford Turnpike, North  
Haven, CT 06473 (US). LANZ, Elizabeth, M. [US/US];  
111 Park Street, No. 54, New Haven, CT 06511 (US).  
NANDABALAN, Krishnan [IN/US]; 228 Village Pond  
Road, Guilford, CT 06437 (US). STEPHENS, Joel,
- (81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ,  
DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR,  
HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,  
LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ,  
NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM,  
TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG,  
CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— Without international search report and to be republished  
upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.

WO 01/11010 A2

(54) Title: DRUG TARGET ISOGENES: POLYMORPHISMS IN THE IMMUNOGLOBULIN E RECEPTOR I ALPHA SUB-  
UNIT GENE

(57) Abstract: Polynucleotides comprising one or more of 22 novel single nucleotide polymorphisms in the human Immunoglobu-  
E Receptor I Alpha Subunit (IGERA) gene are described. Compositions and methods for detecting one or more of these poly-  
phisms are also disclosed. In addition, various genotypes and haplotypes for IGERA gene that exist in the population are desc-

DRUG TARGET ISOGENES:  
POLYMORPHISMS IN THE IMMUNOGLOBULIN E RECEPTOR I ALPHA SUBUNIT  
GENE

## 5 RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Patent Application Serial No. 60/147,860 filed August 9, 1999 and entitled "Receptor Isogenes: Polymorphisms in the Immunoglobulin E Receptor I Alpha Subunit Gene".

## 10 FIELD OF THE INVENTION

This invention relates to variation in genes that encode pharmaceutically important proteins. In particular, this invention provides genetic variants of the human Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene and methods for identifying which variant(s) of this gene is/are possessed by an individual.

15

## BACKGROUND OF THE INVENTION

Current methods for identifying pharmaceuticals to treat disease often start by identifying, cloning, and expressing an important target protein related to the disease. A determination of whether an agonist or antagonist is needed to produce an effect that may benefit a patient with the disease is then made. Then, vast numbers of compounds are screened against the target protein to find new potential drugs. The desired outcome of this process is a drug that is specific for the target, thereby reducing the incidence of the undesired side effects usually caused by a compound's activity at non-intended targets.

What this approach fails to consider, however, is that natural variability exists in any and every population with respect to a particular protein. A target protein currently used to screen drugs typically is expressed by a gene cloned from an individual who was arbitrarily selected. However, the nucleotide sequence of a particular gene may vary tremendously among individuals. Subtle alteration(s) in the primary nucleotide sequence of a gene encoding a target protein may be manifested as significant variation in expression of or in the structure and/or function of the protein. Such alterations may explain the relatively high degree of uncertainty inherent in treatment of individuals with drugs whose design is based upon a single representative example of the target. For example, it is well-established that some classes of drugs frequently have lower efficacy in some individuals than others, which means such individuals and their physicians must weigh the possible benefit of a larger dosage against a greater risk of side effects. In addition, variable information on the biological function or effects of a particular protein may be due to different scientists unknowingly studying different isoforms of the gene encoding the protein. Thus, information on the type and frequency of genomic variation that exists for pharmaceutically important proteins would be useful.

The organization of single nucleotide variations (polymorphisms) in the primary sequence of a

gene into one of the limited number of combinations that exist as units of inheritance is termed a haplotype. Each haplotype therefore contains significantly more information than individual unorganized polymorphisms. Haplotypes provide an accurate measurement of the genomic variation in the two chromosomes of an individual.

5 It is well-established that many diseases are associated with specific variations in gene sequences. However while there are examples in which individual polymorphisms act as genetic markers for a particular phenotype, in other cases an individual polymorphism may be found in a variety of genomic backgrounds and therefore shows no definitive coupling between the polymorphism and the causative site for the phenotype (Clark AG et al. 1998 *Am J Hum Genet* 63:595-612; Ulbrecht  
10 M et al. 2000 *Am J Respir Crit Care Med* 161: 469-74). In addition, the marker may be predictive in some populations, but not in other populations (Clark AG et al. 1998 *supra*). In these instances, a haplotype will provide a superior genetic marker for the phenotype (Clark AG et al. 1998 *supra*; Ulbrecht M et al. 2000, *supra*; Ruaño G & Stephens JC *Gen Eng News* 19 (21), December 1999).

Analysis of the association between each observed haplotype and a particular phenotype  
15 permits ranking of each haplotype by its statistical power of prediction for the phenotype. Haplotypes found to be strongly associated with the phenotype can then have that positive association confirmed by alternative methods to minimize false positives. For a gene suspected to be associated with a particular phenotype, if no observed haplotypes for that gene show association with the phenotype of interest, then it may be inferred that variation in the gene has little, if any, involvement with that phenotype (Ruaño &  
20 Stephens 1999, *supra*). Thus, information on the observed haplotypes and their frequency of occurrence in various population groups will be useful in a variety of research and clinical applications.

One possible drug target involved in immune response is the Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene or its encoded product. The high affinity IgE receptor (IgERI) belongs to the family of antibody Fc receptors that play an important role in the immune response by coupling the  
25 specificity of secreted antibodies to a variety of cells of the immune system. Fc receptors initiate immune system reactions in normal immunity, allergies, antibody-mediated tumor recognition, and autoimmune diseases such as arthritis. The high affinity IgE receptor (IgERI) mediates IgE-dependent peripheral and systemic anaphylaxis, regulates IgE metabolism, and plays a role in the growth and differentiation of various cells of the immune system.

30 The IgERI initiates the immediate hypersensitivity response from mast cells and basophils, and evidence indicates this receptor is involved in antiparasitic reactions from platelet and eosinophils, and in antigen delivery to dendritic cells for MHC class II presentation pathways activating T cells. Moreover, IgERI exerts a regulatory effect on IgE production, as well as differentiation and growth of mast cell and B-lymphocytes. Stimulation of IgERI initiates a cascade of events resulting in a number  
35 of cellular events. Mast cells release inflammatory mediators, such as histamine. Cytokines are released, particularly interleukin 4 (IL-4), which is critical in the B-cell switching and IgE synthesis pathways, as well as a feed-back up-regulation of IgERI synthesis. Expression and functions of other

mast cell surface receptors, such as CD40, involved in immune cell growth and differentiation, as well as IgE metabolism, are induced. Other factors whose expression and/or secretion are regulated by IgERI include, interleukin 6 (IL-6), tissue necrosis factor alpha (TNF $\alpha$ ), RANTES, and serotonin, among others.

- 5 IgERI is a tetrameric transmembrane protein existing consisting of an alpha, beta, and two disulfide-bonded gamma polypeptides. The alpha subunit, IGER $\alpha$ , binds IgE with high affinity (Kd ~109-1010M) and can be secreted as a soluble IgE-binding fragment. The gamma subunit, IgERI $\gamma$ , mediates receptor assembly and signal transduction, and is a common component of other Fc receptors, including the high-affinity and low-affinity IgG receptors, and the TCR/CD3 T-cell receptor complex.
- 10 The role of the beta subunit, IgERI $\beta$ , is more enigmatic, although it is also involved in signal transduction and receptor autophosphorylation. IgERI $\beta$  is essential for full activation of mast cells for the allergic response and is an amplifier of signaling from the gamma subunit.

- IGERA consists of a C-terminal cytoplasmic tail, a single transmembrane region, and an N-terminal extracellular region divided into two large immunoglobulin (Ig) domains. The Ig domains are each 85 amino acids in length, and are bent at an acute angle to form a convex binding site for IgE. The second domain has a prominent loop that projects above the domain and is a site of interaction with IgE.
- 15 IgERI $\beta$  is a four transmembrane protein with N-terminal and C-terminal cytoplasmic tails. The N-terminal cytoplasmic domain interacts with the cytoplasmic domains of the IgERI $\gamma$  subunits. The C-terminal cytoplasmic tail of IgERI $\beta$  associates with the cytoplasmic tail of the alpha subunit. IgERI $\gamma$  has a short extracellular N-terminal tail, a single transmembrane region, and a C-terminal cytoplasmic domain.
- 20

- Both IgERI $\beta$  and IgERI $\gamma$  have an immunoreceptor tyrosine activation motif (ITAM) in their cytoplasmic domains. The IgERI $\beta$  ITAM appears in the C-terminal cytoplasmic domain. Evidence suggests that the two ITAM domains act synergistically, associating with specific protein tyrosine kinases that are capable of triggering cell activation via protein-tyrosine phosphorylation. Receptor subunit cross-linking activates the src kinase, Lyn, associated with the IgERI $\beta$  ITAM, in turn phosphorylating two tyrosine residues in the ITAM. This event activates the src kinase, Syk, associated with the IgERI $\gamma$  ITAM, phosphorylating the ITAM tyrosines in that subunit. Deletion of the C-terminal cytoplasmic domain of IgERI $\beta$ , containing the Lyn ITAM, results in an inactive receptor. Mutation of the either or both tyrosines in the IgERI $\beta$  ITAM results in non-phosphorylation of IgERI $\beta$  and IgERI $\gamma$  tyrosines.
- 25
- 30

- The gene for the alpha subunit of the high-affinity IgE receptor is located on human chromosome band 1q23, along with the gene for the gamma subunit (Tepler et al., *Am. J. Hum. Genet.* 45: 761-765, 1989; Le Coniat et al., *Immunogenetics* 32: 183-186, 1990). The IGERA gene spans approximately 5900 base pairs (bp) of genomic DNA and consists of five exons encoding 257 amino acids (Kochan et al., *Nucl. Acids Res.* 16:3584, 1988; Shimizu et al., *Proc. Natl. Acad. Sci., USA*
- 35

85:1907-1911, 1988). Reference sequences for the IGERA gene (GenBank Accession No. L14075; SEQ ID NO:1), coding sequence and protein are shown in Figs. 1, 2, and 3 respectively. Significant features reported for the IGERA gene and its encoded protein include: enhancer binding motifs at nucleotide positions 1184-1189 and 1203-1209 for Ets- and GATA-family transcription factors; 29 bp of 5' untranslated region in the first exon; the ATG initiation codon at nucleotide position 1287; a first extracellular domain located between amino acids 1-85; a second extracellular domain located between amino acids 86-170; a transmembrane region between amino acids 205-226; and a C-terminal cytoplasmic region between amino acids 227-257.

Interest in discovering polymorphisms in genes encoding subunits of IgERI arises from the role played by IgE in atopy. Atopy is a common familial disorder caused by genetic and environmental factors. Atopy is characterized by exaggerated T- helper cell type II lymphocyte responses to common allergens, such as pollens and dust mites, and included sustained, enhanced production of IgE. Allergy, asthma, rhinitis, and eczema are atopic hypersensitivity diseases. IgE binds to the high affinity IgE receptor presented on mucosal mast cells and basophils. IgE binding of allergens activates the receptor and initiates a cascade, leading to cellular release of inflammatory mediators. Dysregulation of the normal immediate hypersensitivity response results in abnormally high and sustained IgE serum levels, which leads to mucosal inflammation. Atopy is detected by elevated total serum IgE levels, positive skin prick tests to common allergens, and specific serum IgE against these allergens. All three have been strongly correlated with each other and the presence of the symptoms of allergic reaction such as wheezing, coughing, sneezing, and nasal blockage.

Approximately 20% of the world population is affected by allergies, with over 50% of western populations testing positive to skin prick tests of one or more common allergens. Up to 10% of children suffer from atopic asthma, accounting for approximately one-third of pediatric emergency room visits in the United States. While a single genetic determinant is unlikely to be the causative factor in asthma, allergy, or other atopic diseases, therapeutics aimed at the obligatory binding of IgE to IgERI for initiation of the allergic response could provide a single treatment for the various manifestations of atopic hypersensitivity.

Few published studies have been performed to identify polymorphisms at the IGERA locus. One known polymorphism at the IGERA locus consists of an RsaI restriction fragment length polymorphism (RFLP) detected in genomic DNA using a cDNA probe (Tepler et al., *supra*). Fragments of 1.8, 0.6, and 0.3 kilobase pairs (kb) were detected in all individuals tested, with a variant fragment of 2.8 kb detected at a 40% frequency. The location of the polymorphic RsaI site within the gene is unknown, but is mostly likely intronic.

Because of the potential for polymorphisms in the IGERA gene to affect the expression and function of the encoded protein, it would be useful to determine whether polymorphisms exist in the IGERA gene, as well as how such polymorphisms are combined in different copies of the gene. Such information would be useful for studying the biological function of IGERA as well as in identifying

drugs targeting this protein for the treatment of disorders related to its abnormal expression or function.

#### SUMMARY OF THE INVENTION

Accordingly, the inventors herein have discovered 22 novel polymorphic sites in the IGERA gene. These polymorphic sites (PS) correspond to the following nucleotide positions in the indicated GenBank Accession Number: 872 (PS1), 943 (PS2), 1192 (PS3), 1199 (PS4), 1363 (PS5), 1754 (PS6), 1760 (PS7), 1896 (PS8), 2708 (PS9), 3024 (PS10), 3075 (PS11), 3220 (PS12), 3286 (PS13), 3330 (PS14), 4838 (PS15), 5108 (PS16), 5285 (PS17), 5363 (PS18), 6821 (PS19), 6911 (PS20), 6936 (PS21) and 7000 (PS22) in L14075. The polymorphisms at these sites are thymine or guanine at PS1, thymine or cytosine at PS2, thymine or cytosine at PS3, adenine or thymine at PS4, cytosine or adenine at PS5, thymine or cytosine at PS6, cytosine or adenine at PS7, cytosine or thymine at PS8, adenine or guanine at PS9, adenine or guanine at PS10, guanine or adenine at PS11, thymine or cytosine at PS12, guanine or adenine at PS13, guanine or adenine at PS14, guanine or adenine at PS15, cytosine or thymine at PS16, cytosine or thymine at PS17, thymine or cytosine at PS18, cytosine or adenine at PS19, thymine or cytosine at PS20, adenine or guanine at PS21 and guanine or adenine at PS22. In addition, the inventors have determined the identity of the alternative nucleotides present at these sites in a human reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: African descent, Asian, Caucasian and Hispanic/Latino. It is believed that IGERA-encoding polynucleotides containing one or more of the novel polymorphic sites reported herein will be useful in studying the expression and biological function of IGERA, as well as in developing drugs targeting this protein. In addition, information on the combinations of polymorphisms in the IGERA gene may have diagnostic and forensic applications.

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the IGERA gene or a fragment thereof. The reference sequence comprises SEQ ID NO:1 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, adenine at PS7, thymine at PS8, guanine at PS9, guanine at PS10, adenine at PS11, cytosine at PS12, adenine at PS13, adenine at PS14, adenine at PS15, thymine at PS16, thymine at PS17, cytosine at PS18, adenine at PS19, cytosine at PS20, guanine at PS21 and adenine at PS22. A particularly preferred polymorphic variant is a naturally-occurring isoform (also referred to herein as an "isogene") of the IGERA gene. A IGERA isogene of the invention comprises guanine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, adenine at PS7, thymine at PS8, guanine at PS9, guanine at PS10, adenine at PS11, cytosine at PS12, adenine at PS13, adenine at PS14, adenine at PS15, thymine at PS16, thymine at PS17, cytosine at PS18, adenine at PS19, cytosine at PS20, guanine at PS21 and adenine at PS22. The invention also provides a collection of IGERA isogenes, referred to herein as a IGERA genome anthology.

5 A IGERA isogene may be defined by the combination and order of these polymorphisms in the isogene, which is referred to herein as a IGERA haplotype. Thus, the invention also provides data on the number of different IGERA haplotypes found in the above four population groups. This haplotype data is useful in methods for deriving a IGERA haplotype from an individual's genotype for the IGERA gene and for determining an association between a IGERA haplotype and a particular trait.

In another embodiment, the invention provides a polynucleotide comprising a polymorphic variant of a reference sequence for a IGERA cDNA or a fragment thereof. The reference sequence comprises SEQ ID NO:2 (Fig. 2) and the polymorphic cDNA comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 251, adenine at a position corresponding to nucleotide 302, thymine at a position corresponding to nucleotide 530 and adenine at a position corresponding to nucleotide 741.

Polynucleotides complementary to these IGERA genomic and cDNA variants are also provided by the invention.

15 In other embodiments, the invention provides a recombinant expression vector comprising one of the polymorphic genomic variants operably linked to expression regulatory elements as well as a recombinant host cell transformed or transfected with the expression vector. The recombinant vector and host cell may be used to express IGERA for protein structure analysis and drug binding studies.

In yet another embodiment, the invention provides a polypeptide comprising a polymorphic variant of a reference amino acid sequence for the IGERA protein. The reference amino acid sequence comprises SEQ ID NO:3 (Fig. 3) and the polymorphic variant comprises at least one variant amino acid selected from the group consisting of arginine at a position corresponding to amino acid position 84, asparagine at a position corresponding to amino acid position 101, methionine at a position corresponding to amino acid position 177 and lysine at a position corresponding to amino acid position 247. A polymorphic variant of IGERA is useful in studying the effect of the variation on the biological activity of IGERA as well as studying the binding affinity of candidate drugs targeting IGERA for the treatment of immune response.

The present invention also provides antibodies that recognize and bind to the above polymorphic IGERA protein variant. Such antibodies can be utilized in a variety of diagnostic and prognostic formats and therapeutic methods.

30 In other embodiments, the invention provides methods, compositions, and kits for haplotyping and/or genotyping the IGERA gene in an individual. The methods involve identifying the nucleotide or nucleotide pair present at one or more polymorphic sites selected from PS1-22 in one or both copies of the IGERA gene from the individual. The compositions contain oligonucleotide probes and primers designed to specifically hybridize to one or more target regions containing, or that are adjacent to, a polymorphic site. The methods and compositions for establishing the genotype or haplotype of an individual at the novel polymorphic sites described herein are useful for studying the effect of the polymorphisms in the etiology of diseases affected by the expression and function of the IGERA

protein, studying the efficacy of drugs targeting IGERA, predicting individual susceptibility to diseases affected by the expression and function of the IGERA protein and predicting individual responsiveness to drugs targeting IGERA.

In yet another embodiment, the invention provides a method for identifying an association  
5 between a genotype or haplotype and a trait. In preferred embodiments, the trait is susceptibility to a disease, severity of a disease, the staging of a disease or response to a drug. Such methods have applicability in developing diagnostic tests and therapeutic treatments for immune response.

The present invention also provides transgenic animals comprising one of the IGERA genomic polymorphic variants described herein and methods for producing such animals. The transgenic  
10 animals are useful for studying expression of the IGERA isogenes *in vivo*, for *in vivo* screening and testing of drugs targeted against IGERA protein, and for testing the efficacy of therapeutic agents and compounds for immune response in a biological system.

The present invention also provides a computer system for storing and displaying polymorphism data determined for the IGERA gene. The computer system comprises a computer  
15 processing unit; a display; and a database containing the polymorphism data. The polymorphism data includes the polymorphisms, the genotypes and the haplotypes identified for the IGERA gene in a reference population. In a preferred embodiment, the computer system is capable of producing a display showing IGERA haplotypes organized according to their evolutionary relationships.

## 20 BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 illustrates a reference sequence for the IGERA gene (Genbank Version Number L14075; contiguous lines; SEQ ID NO:1), with the start and stop positions of each region of coding sequence indicated below the sequence by the numbers within the brackets and the polymorphic sites and polymorphisms identified by Applicants in a reference population indicated by the variant  
25 nucleotide positioned below the polymorphic site in the sequence.

Figure 2 illustrates a reference sequence for the IGERA coding sequence (contiguous lines; SEQ ID NO:2), with the polymorphic sites and polymorphisms identified by Applicants in a reference population indicated by the variant nucleotide positioned below the polymorphic site in the sequence.

Figure 3 illustrates a reference sequence for the IGERA protein (contiguous lines; SEQ ID  
30 NO:3), with the variant amino acids caused by the polymorphisms of Fig. 2 positioned below the polymorphic site in the sequence. Any exclamation points (!) presented below the reference sequence represent a termination codon introduced by a polymorphism of Figure 2.

## DESCRIPTION OF THE PREFERRED EMBODIMENTS

35 The present invention is based on the discovery of novel variants of the IGERA gene. As described in more detail below, the inventors herein discovered 22 novel polymorphic sites by characterizing the IGERA gene found in genomic DNAs isolated from an Index Repository that



contains immortalized cell lines from one chimpanzee and 93 human individuals. The human individuals included a reference population of 79 unrelated individuals self-identified as belonging to one of four major population groups: Caucasian (22 individuals), African descent (20 individuals) Asian (20 individuals) Hispanic/Latino (17 individuals). To the extent possible, the members of this reference population were organized into population subgroups by the self-identified ethnogeographic origin of their four grandparents as shown in Table 1 below.

Table 1. Population Groups in the Index Repository

Population Group	Population Subgroup	No. of Individuals
African descent		20
	Sierra Leone	1
Asian		20
	Burma	1
	China	3
	Japan	6
	Korea	1
	Philippines	5
	Vietnam	4
Caucasian		22
	British Isles	3
	British Isles/Central	4
	British Isles/Eastern	1
	Central/Eastern	1
	Eastern	3
	Central/Mediterranean	1
	Mediterranean	2
	Scandinavian	2
Hispanic/Latino		17
	Caribbean	7
	Caribbean (Spanish Descent)	2
	Central American (Spanish Descent)	1
	Mexican American	4
	South American (Spanish Descent)	3

In addition, the Index Repository contains three unrelated indigenous American Indians (one from each of North, Central and South America), one three-generation Caucasian family (from the CEPH Utah cohort) and one two-generation African-American family.

Using the IGERA genotypes identified in the Index Repository and the methodology described in the Examples below, the inventors herein also determined the haplotypes found on each chromosome for most human members of this repository. The IGERA genotypes and haplotypes found in the repository include those shown in Tables 4 and 5, respectively. The polymorphism and haplotype data disclosed herein are useful for studying population diversity, anthropological lineage, the significance of diversity and lineage at the phenotypic level, paternity testing, forensic applications, and for identifying associations between the IGERA genetic variation and a trait such as level of drug response or susceptibility to disease.

In the context of this disclosure, the following terms shall be defined as follows unless

otherwise indicated:

**Allele** - A particular form of a genetic locus, distinguished from other forms by its particular nucleotide sequence.

**Candidate Gene** - A gene which is hypothesized to be responsible for a disease, condition, or the response to a treatment, or to be correlated with one of these.

**Gene** - A segment of DNA that contains all the information for the regulated biosynthesis of an RNA product, including promoters, exons, introns, and other untranslated regions that control expression.

**Genotype** - An unphased 5' to 3' sequence of nucleotide pair(s) found at one or more polymorphic sites in a locus on a pair of homologous chromosomes in an individual. As used herein, genotype includes a full-genotype and/or a sub-genotype as described below.

**Full-genotype** - The unphased 5' to 3' sequence of nucleotide pairs found at all known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

**Sub-genotype** - The unphased 5' to 3' sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a pair of homologous chromosomes in a single individual.

**Genotyping** - A process for determining a genotype of an individual.

**Haplotype** - A 5' to 3' sequence of nucleotides found at one or more polymorphic sites in a locus on a single chromosome from a single individual. As used herein, haplotype includes a full-haplotype and/or a sub-haplotype as described below.

**Full-haplotype** - The 5' to 3' sequence of nucleotides found at all known polymorphic sites in a locus on a single chromosome from a single individual.

**Sub-haplotype** - The 5' to 3' sequence of nucleotides seen at a subset of the known polymorphic sites in a locus on a single chromosome from a single individual.

**Haplotype pair** - The two haplotypes found for a locus in a single individual.

**Haplotyping** - A process for determining one or more haplotypes in an individual and includes use of family pedigrees, molecular techniques and/or statistical inference.

**Haplotype data** - Information concerning one or more of the following for a specific gene: a listing of the haplotype pairs in each individual in a population; a listing of the different haplotypes in a population; frequency of each haplotype in that or other populations, and any known associations between one or more haplotypes and a trait.

**Isoform** - A particular form of a gene, mRNA, cDNA or the protein encoded thereby, distinguished from other forms by its particular sequence and/or structure.

**Isogene** - One of the isoforms of a gene found in a population. An isogene contains all of the polymorphisms present in the particular isoform of the gene.

**Isolated** - As applied to a biological molecule such as RNA, DNA, oligonucleotide, or protein, isolated means the molecule is substantially free of other biological molecules such as nucleic acids, proteins, lipids, carbohydrates, or other material such as cellular debris and growth media. Generally,

the term "isolated" is not intended to refer to a complete absence of such material or to absence of water, buffers, or salts, unless they are present in amounts that substantially interfere with the methods of the present invention.

**Locus** - A location on a chromosome or DNA molecule corresponding to a gene or a physical or phenotypic feature.

**Naturally-occurring** - A term used to designate that the object it is applied to, e.g., naturally-occurring polynucleotide or polypeptide, can be isolated from a source in nature and which has not been intentionally modified by man.

**Nucleotide pair** - The nucleotides found at a polymorphic site on the two copies of a chromosome from an individual.

**Phased** - As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, phased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is known.

**Polymorphic site (PS)** - A position within a locus at which at least two alternative sequences are found in a population, the most frequent of which has a frequency of no more than 99%.

**Polymorphic variant** - A gene, mRNA, cDNA, polypeptide or peptide whose nucleotide or amino acid sequence varies from a reference sequence due to the presence of a polymorphism in the gene.

**Polymorphism** - The sequence variation observed in an individual at a polymorphic site. Polymorphisms include nucleotide substitutions, insertions, deletions and microsatellites and may, but need not, result in detectable differences in gene expression or protein function.

**Polymorphism data** - Information concerning one or more of the following for a specific gene: location of polymorphic sites; sequence variation at those sites; frequency of polymorphisms in one or more populations; the different genotypes and/or haplotypes determined for the gene; frequency of one or more of these genotypes and/or haplotypes in one or more populations; any known association(s) between a trait and a genotype or a haplotype for the gene.

**Polymorphism Database** - A collection of polymorphism data arranged in a systematic or methodical way and capable of being individually accessed by electronic or other means.

**Polynucleotide** - A nucleic acid molecule comprised of single-stranded RNA or DNA or comprised of complementary, double-stranded DNA.

**Population Group** - A group of individuals sharing a common ethnogeographic origin.

**Reference Population** - A group of subjects or individuals who are predicted to be representative of the genetic variation found in the general population. Typically, the reference population represents the genetic variation in the population at a certainty level of at least 85%, preferably at least 90%, more preferably at least 95% and even more preferably at least 99%.

**Single Nucleotide Polymorphism (SNP)** - Typically, the specific pair of nucleotides observed at a single polymorphic site. In rare cases, three or four nucleotides may be found.

**Subject** – A human individual whose genotypes or haplotypes or response to treatment or disease state are to be determined.

**Treatment** - A stimulus administered internally or externally to a subject.

**Unphased** – As applied to a sequence of nucleotide pairs for two or more polymorphic sites in a locus, unphased means the combination of nucleotides present at those polymorphic sites on a single copy of the locus is not known.

The inventors herein have discovered 22 novel polymorphic sites in the IGERA gene. The polymorphic sites identified by the inventors are referred to as PS1-22 to designate the order in which they are located in the gene (see Table 3 below).

Thus, in one embodiment, the invention provides an isolated polynucleotide comprising a polymorphic variant of the IGERA gene or a fragment of the gene which contains at least one of the novel polymorphic sites described herein. The nucleotide sequence of a variant IGERA gene is identical to the reference genomic sequence for those portions of the gene examined, as described in the Examples below, except that it comprises a different nucleotide at one or more of the novel polymorphic sites PS1-22. Similarly, the nucleotide sequence of a variant fragment of the IGERA gene is identical to the corresponding portion of the reference sequence except for having a different nucleotide at one or more of the novel polymorphic sites described herein. Thus, the invention specifically does not include polynucleotides comprising a nucleotide sequence identical to the reference sequence (or other reported IGERA sequences) or to portions of the reference sequence (or other reported IGERA sequences), except for genotyping oligonucleotides as described below.

The location of a polymorphism in a variant gene or fragment is identified by aligning its sequence against SEQ ID NO:1. The polymorphism is selected from the group consisting of guanine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, adenine at PS7, thymine at PS8, guanine at PS9, guanine at PS10, adenine at PS11, cytosine at PS12, adenine at PS13, adenine at PS14, adenine at PS15, thymine at PS16, thymine at PS17, cytosine at PS18, adenine at PS19, cytosine at PS20, guanine at PS21 and adenine at PS22. In a preferred embodiment, the polymorphic variant comprises a naturally-occurring isogene of the IGERA gene which is defined by any one of haplotypes 1-20 shown in Table 5 below.

Polymorphic variants of the invention may be prepared by isolating a clone containing the IGERA gene from a human genomic library. The clone may be sequenced to determine the identity of the nucleotides at the polymorphic sites described herein. Any particular variant claimed herein could be prepared from this clone by performing *in vitro* mutagenesis using procedures well-known in the art.

IGERA isogenes may be isolated using any method that allows separation of the two "copies" of the IGERA gene present in an individual, which, as readily understood by the skilled artisan, may be the same allele or different alleles. Separation methods include targeted *in vivo* cloning (TIVC) in yeast as described in WO 98/01573, U.S. Patent No. 5,866,404, and copending U.S. application Serial No. 08/987,966. Another method, which is described in copending U.S. Application Serial No. 08/987,966,

uses an allele specific oligonucleotide in combination with primer extension and exonuclease degradation to generate hemizygous DNA targets. Yet other methods are single molecule dilution (SMD) as described in Ruaño et al., Proc. Natl. Acad. Sci. 87:6296-6300, 1990; and allele specific PCR (Ruaño et al., 17 Nucleic Acids. Res. 8392, 1989; Ruaño et al., 19 Nucleic Acids Res. 6877-6882, 1991; Michalatos-Beloin et al., 24 Nucleic Acids Res. 4841-4843, 1996).

The invention also provides IGERA genome anthologies, which are collections of IGERA isogenes found in a given population. The population may be any group of at least two individuals, including but not limited to a reference population, a population group, a family population, a clinical population, and a same sex population. A IGERA genome anthology may comprise individual IGERA isogenes stored in separate containers such as microtest tubes, separate wells of a microtitre plate and the like. Alternatively, two or more groups of the IGERA isogenes in the anthology may be stored in separate containers. Individual isogenes or groups of isogenes in a genome anthology may be stored in any convenient and stable form, including but not limited to in buffered solutions, as DNA precipitates, freeze-dried preparations and the like. A preferred IGERA genome anthology of the invention comprises a set of isogenes defined by the haplotypes shown in Table 5 below.

An isolated polynucleotide containing a polymorphic variant nucleotide sequence of the invention may be operably linked to one or more expression regulatory elements in a recombinant expression vector capable of being propagated and expressing the encoded IGERA protein in a prokaryotic or a eukaryotic host cell. Examples of expression regulatory elements which may be used include, but are not limited to, the lac system, operator and promoter regions of phage lambda, yeast promoters, and promoters derived from vaccinia virus, adenovirus, retroviruses, or SV40. Other regulatory elements include, but are not limited to, appropriate leader sequences, termination codons, polyadenylation signals, and other sequences required for the appropriate transcription and subsequent translation of the nucleic acid sequence in a given host cell. Of course, the correct combinations of expression regulatory elements will depend on the host system used. In addition, it is understood that the expression vector contains any additional elements necessary for its transfer to and subsequent replication in the host cell. Examples of such elements include, but are not limited to, origins of replication and selectable markers. Such expression vectors are commercially available or are readily constructed using methods known to those in the art (e.g., F. Ausubel et al., 1987, in "Current Protocols in Molecular Biology", John Wiley and Sons, New York, New York). Host cells which may be used to express the variant IGERA sequences of the invention include, but are not limited to, eukaryotic and mammalian cells, such as animal, plant, insect and yeast cells, and prokaryotic cells, such as E. coli, or algal cells as known in the art. The recombinant expression vector may be introduced into the host cell using any method known to those in the art including, but not limited to, microinjection, electroporation, particle bombardment, transduction, and transfection using DEAE-dextran, lipofection, or calcium phosphate (see e.g., Sambrook et al. (1989) in "Molecular Cloning. A Laboratory Manual", Cold Spring Harbor Press, Plainview, New York). In a preferred aspect, eukaryotic expression vectors

that function in eukaryotic cells, and preferably mammalian cells, are used. Non-limiting examples of such vectors include vaccinia virus vectors, adenovirus vectors, herpes virus vectors, and baculovirus transfer vectors. Preferred eukaryotic cell lines include COS cells, CHO cells, HeLa cells, NIH/3T3 cells, and embryonic stem cells (Thomson, J. A. et al., 1998 Science 282:1145-1147). Particularly preferred host cells are mammalian cells.

As will be readily recognized by the skilled artisan, expression of polymorphic variants of the IGERA gene will produce IGERA mRNAs varying from each other at any polymorphic site retained in the spliced and processed mRNA molecules. These mRNAs can be used for the preparation of a IGERA cDNA comprising a nucleotide sequence which is a polymorphic variant of the IGERA reference coding sequence shown in Figure 2. Thus, the invention also provides IGERA mRNAs and corresponding cDNAs which comprise a nucleotide sequence that is identical to SEQ ID NO:2 (Fig. 2), or its corresponding RNA sequence, except for having one or more polymorphisms selected from the group consisting of guanine at a position corresponding to nucleotide 251, adenine at a position corresponding to nucleotide 302, thymine at a position corresponding to nucleotide 530 and adenine at a position corresponding to nucleotide 741. Fragments of these variant mRNAs and cDNAs are included in the scope of the invention, provided they contain the novel polymorphisms described herein. The invention specifically excludes polynucleotides identical to previously identified and characterized IGERA cDNAs and fragments thereof. Polynucleotides comprising a variant RNA or DNA sequence may be isolated from a biological sample using well-known molecular biological procedures or may be chemically synthesized.

Genomic and cDNA fragments of the invention comprise at least one novel polymorphic site identified herein and have a length of at least 10 nucleotides and may range up to the full length of the gene. Preferably, a fragment according to the present invention is between 100 and 3000 nucleotides in length, and more preferably between 200 and 2000 nucleotides in length, and most preferably between 500 and 1000 nucleotides in length.

In describing the polymorphic sites identified herein, reference is made to the sense strand of the gene for convenience. However, as recognized by the skilled artisan, nucleic acid molecules containing the IGERA gene may be complementary double stranded molecules and thus reference to a particular site on the sense strand refers as well to the corresponding site on the complementary antisense strand. Thus, reference may be made to the same polymorphic site on either strand and an oligonucleotide may be designed to hybridize specifically to either strand at a target region containing the polymorphic site. Thus, the invention also includes single-stranded polynucleotides which are complementary to the sense strand of the IGERA genomic variants described herein.

Polynucleotides comprising a polymorphic gene variant or fragment may be useful for therapeutic purposes. For example, where a patient could benefit from expression, or increased expression, of a particular IGERA protein isoform, an expression vector encoding the isoform may be administered to the patient. The patient may be one who lacks the IGERA isogene encoding that

isoform or may already have at least one copy of that isogene.

In other situations, it may be desirable to decrease or block expression of a particular IGERA isogene. Expression of a IGERA isogene may be turned off by transforming a targeted organ, tissue or cell population with an expression vector that expresses high levels of untranslatable mRNA for the isogene. Alternatively, oligonucleotides directed against the regulatory regions (e.g., promoter, introns, enhancers, 3' untranslated region) of the isogene may block transcription. Oligonucleotides targeting the transcription initiation site, e.g., between positions -10 and +10 from the start site are preferred. Similarly, inhibition of transcription can be achieved using oligonucleotides that base-pair with region(s) of the isogene DNA to form triplex DNA (see e.g., Gee et al. in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt. Kisco, N.Y., 1994). Antisense oligonucleotides may also be designed to block translation of IGERA mRNA transcribed from a particular isogene. It is also contemplated that ribozymes may be designed that can catalyze the specific cleavage of IGERA mRNA transcribed from a particular isogene.

The oligonucleotides may be delivered to a target cell or tissue by expression from a vector introduced into the cell or tissue *in vivo* or *ex vivo*. Alternatively, the oligonucleotides may be formulated as a pharmaceutical composition for administration to the patient. Oligoribonucleotides and/or oligodeoxynucleotides intended for use as antisense oligonucleotides may be modified to increase stability and half-life. Possible modifications include, but are not limited to phosphorothioate or 2' O-methyl linkages, and the inclusion of nontraditional bases such as inosine and queosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous nucleases.

The invention also provides an isolated polypeptide comprising a polymorphic variant of the reference IGERA amino acid sequence shown in Figure 3. The location of a variant amino acid in a IGERA polypeptide or fragment of the invention is identified by aligning its sequence against SEQ ID NO:3. A IGERA protein variant of the invention comprises an amino acid sequence identical to SEQ ID NO: 3 except for having one or more variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 84, asparagine at a position corresponding to amino acid position 101, methionine at a position corresponding to amino acid position 177 and lysine at a position corresponding to amino acid position 247. The invention specifically excludes amino acid sequences identical to those previously identified for IGERA, including SEQ ID NO: 3, and previously described fragments thereof. IGERA protein variants included within the invention comprise all amino acid sequences based on SEQ ID NO: 3 and having the combination of amino acid variations described in Table 2 below. In preferred embodiments, a IGERA protein variant of the invention is encoded by an isogene defined by one of the observed haplotypes shown in Table 5.

Table 2. Novel Polymorphic Variant of IGERA

	Polymorphic Variant Number	Amino Acid Position and Identities			
		84	101	177	247
5	1	K	S	T	K
	2	K	S	M	N
	3	K	S	M	K
	4	K	N	T	N
10	5	K	N	T	K
	6	K	N	M	N
	7	K	N	M	K
	8	R	S	T	N
	9	R	S	T	K
15	10	R	S	M	N
	11	R	S	M	K
	12	R	N	T	N
	13	R	N	T	K
	14	R	N	M	N
20	15	R	N	M	K

The invention also includes IGERA peptide variants, which are any fragments of a IGERA protein variant that contains one or more of the amino acid variations shown in Table 2. A IGERA peptide variant is at least 6 amino acids in length and is preferably any number between 6 and 30 amino acids long, more preferably between 10 and 25, and most preferably between 15 and 20 amino acids long. Such IGERA peptide variants may be useful as antigens to generate antibodies specific for one of the above IGERA isoforms. In addition, the IGERA peptide variants may be useful in drug screening assays.

A IGERA variant protein or peptide of the invention may be prepared by chemical synthesis or by expressing one of the variant IGERA genomic and cDNA sequences as described above. Alternatively, the IGERA protein variant may be isolated from a biological sample of an individual having a IGERA isogene which encodes the variant protein. Where the sample contains two different IGERA isoforms (i.e., the individual has different IGERA isogenes), a particular IGERA isoform of the invention can be isolated by immunoaffinity chromatography using an antibody which specifically binds to that particular IGERA isoform but does not bind to the other IGERA isoform.

The expressed or isolated IGERA protein may be detected by methods known in the art, including Coomassie blue staining, silver staining, and Western blot analysis using antibodies specific for the isoform of the IGERA protein as discussed further below. IGERA variant proteins can be purified by standard protein purification procedures known in the art, including differential precipitation, molecular sieve chromatography, ion-exchange chromatography, isoelectric focusing, gel electrophoresis, affinity and immunoaffinity chromatography and the like. (Ausubel et. al., 1987, In Current Protocols in Molecular Biology John Wiley and Sons, New York, New York). In the case of immunoaffinity chromatography, antibodies specific for a particular polymorphic variant may be used.

A polymorphic variant IGERA gene of the invention may also be fused in frame with a



heterologous sequence to encode a chimeric IGERA protein. The non-IGERA portion of the chimeric protein may be recognized by a commercially available antibody. In addition, the chimeric protein may also be engineered to contain a cleavage site located between the IGERA and non-IGERA portions so that the IGERA protein may be cleaved and purified away from the non-IGERA portion.

5           An additional embodiment of the invention relates to using a novel IGERA protein isoform in any of a variety of drug screening assays. Such screening assays may be performed to identify agents that bind specifically to all known IGERA protein isoforms or to only a subset of one or more of these isoforms. The agents may be from chemical compound libraries, peptide libraries and the like. The IGERA protein or peptide variant may be free in solution or affixed to a solid support. In one  
10           embodiment, high throughput screening of compounds for binding to a IGERA variant may be accomplished using the method described in PCT application WO84/03565, in which large numbers of test compounds are synthesized on a solid substrate, such as plastic pins or some other surface, contacted with the IGERA protein(s) of interest and then washed. Bound IGERA protein(s) are then detected using methods well-known in the art.

15           In another embodiment, a novel IGERA protein isoform may be used in assays to measure the binding affinities of one or more candidate drugs targeting the IGERA protein.

          In another embodiment, the invention provides antibodies specific for and immunoreactive with one or more of the novel IGERA variant proteins described herein. The antibodies may be either monoclonal or polyclonal in origin. The IGERA protein or peptide variant used to generate the  
20           antibodies may be from natural or recombinant sources or produced by chemical synthesis using synthesis techniques known in the art. If the IGERA protein variant is of insufficient size to be antigenic, it may be conjugated, complexed, or otherwise covalently linked to a carrier molecule to enhance the antigenicity of the peptide. Examples of carrier molecules, include, but are not limited to, albumins (e.g., human, bovine, fish, ovine), and keyhole limpet hemocyanin (Basic and Clinical  
25           Immunology, 1991, Eds. D.P. Stites, and A.I. Terr, Appleton and Lange, Norwalk Connecticut, San Mateo, California).

          In one embodiment, an antibody specifically immunoreactive with one of the novel IGERA protein isoforms described herein is administered to an individual to neutralize activity of the IGERA isoform expressed by that individual. The antibody may be formulated as a pharmaceutical  
30           composition which includes a pharmaceutically acceptable carrier.

          Antibodies specific for and immunoreactive with one of the novel IGERA protein isoform described herein may be used to immunoprecipitate the IGERA protein variant from solution as well as react with IGERA protein isoforms on Western or immunoblots of polyacrylamide gels on membrane supports or substrates. In another preferred embodiment, the antibodies will detect IGERA protein  
35           isoforms in paraffin or frozen tissue sections, or in cells which have been fixed or unfixed and prepared on slides, coverslips, or the like, for use in immunocytochemical, immunohistochemical, and

immunofluorescence techniques.

In another embodiment, an antibody specifically immunoreactive with one of the novel IGERA protein variants described herein is used in immunoassays to detect this variant in biological samples. In this method, an antibody of the present invention is contacted with a biological sample and the formation of a complex between the IGERA protein variant and the antibody is detected. As described, suitable immunoassays include radioimmunoassay, Western blot assay, immunofluorescent assay, enzyme linked immunoassay (ELISA), chemiluminescent assay, immunohistochemical assay, immunocytochemical assay, and the like (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Press, New York, New York; Current Protocols in Molecular Biology, 1987, Eds. Ausubel et al., John Wiley and Sons, New York, New York). Standard techniques known in the art for ELISA are described in Methods in Immunodiagnosis, 2nd Ed., Eds. Rose and Bigazzi, John Wiley and Sons, New York 1980; and Campbell et al., 1984, Methods in Immunology, W.A. Benjamin, Inc.). Such assays may be direct, indirect, competitive, or noncompetitive as described in the art (see, e.g., Principles and Practice of Immunoassay, 1991, Eds. Christopher P. Price and David J. Neoman, Stockton Pres, NY, NY; and Oellirich, M., 1984, J. Clin. Chem. Clin. Biochem., 22:895-904). Proteins may be isolated from test specimens and biological samples by conventional methods, as described in Current Protocols in Molecular Biology, supra.

Exemplary antibody molecules for use in the detection and therapy methods of the present invention are intact immunoglobulin molecules, substantially intact immunoglobulin molecules, or those portions of immunoglobulin molecules that contain the antigen binding site. Polyclonal or monoclonal antibodies may be produced by methods conventionally known in the art (e.g., Kohler and Milstein, 1975, Nature, 256:495-497; Campbell Monoclonal Antibody Technology, the Production and Characterization of Rodent and Human Hybridomas, 1985, In: Laboratory Techniques in Biochemistry and Molecular Biology, Eds. Burdon et al., Volume 13, Elsevier Science Publishers, Amsterdam). The antibodies or antigen binding fragments thereof may also be produced by genetic engineering. The technology for expression of both heavy and light chain genes in E. coli is the subject of PCT patent applications, publication number WO 901443, WO 901443 and WO 9014424 and in Huse et al., 1989, Science, 246:1275-1281. The antibodies may also be humanized (e.g., Queen, C. et al. 1989 Proc. Natl. Acad. Sci. 86:10029).

Effect(s) of the polymorphisms identified herein on expression of IGERA may be investigated by preparing recombinant cells and/or organisms, preferably recombinant animals, containing a polymorphic variant of the IGERA gene. As used herein, "expression" includes but is not limited to one or more of the following: transcription of the gene into precursor mRNA; splicing and other processing of the precursor mRNA to produce mature mRNA; mRNA stability; translation of the mature mRNA into IGERA protein (including codon usage and tRNA availability); and glycosylation and/or other modifications of the translation product, if required for proper expression and function.

To prepare a recombinant cell of the invention, the desired IGERA isogene may be introduced into the cell in a vector such that the isogene remains extrachromosomal. In such a situation, the gene will be expressed by the cell from the extrachromosomal location. In a preferred embodiment, the IGERA isogene is introduced into a cell in such a way that it recombines with the endogenous IGERA gene present in the cell. Such recombination requires the occurrence of a double recombination event, thereby resulting in the desired IGERA gene polymorphism. Vectors for the introduction of genes both for recombination and for extrachromosomal maintenance are known in the art, and any suitable vector or vector construct may be used in the invention. Methods such as electroporation, particle bombardment, calcium phosphate co-precipitation and viral transduction for introducing DNA into cells are known in the art; therefore, the choice of method may lie with the competence and preference of the skilled practitioner. Examples of cells into which the IGERA isogene may be introduced include, but are not limited to, continuous culture cells, such as COS, NIH/3T3, and primary or culture cells of the relevant tissue type, i.e., they express the IGERA isogene. Such recombinant cells can be used to compare the biological activities of the different protein variants.

Recombinant organisms, i.e., transgenic animals, expressing a variant IGERA gene are prepared using standard procedures known in the art. Preferably, a construct comprising the variant gene is introduced into a nonhuman animal or an ancestor of the animal at an embryonic stage, i.e., the one-cell stage, or generally not later than about the eight-cell stage. Transgenic animals carrying the constructs of the invention can be made by several methods known to those having skill in the art. One method involves transfecting into the embryo a retrovirus constructed to contain one or more insulator elements, a gene or genes of interest, and other components known to those skilled in the art to provide a complete shuttle vector harboring the insulated gene(s) as a transgene, see e.g., U.S. Patent No. 5,610,053. Another method involves directly injecting a transgene into the embryo. A third method involves the use of embryonic stem cells. Examples of animals into which the IGERA isogenes may be introduced include, but are not limited to, mice, rats, other rodents, and nonhuman primates (see "The Introduction of Foreign Genes into Mice" and the cited references therein, In: Recombinant DNA, Eds. J.D. Watson, M. Gilman, J. Witkowski, and M. Zoller; W.H. Freeman and Company, New York, pages 254-272). Transgenic animals stably expressing a human IGERA isogene and producing human IGERA protein can be used as biological models for studying diseases related to abnormal IGERA expression and/or activity, and for screening and assaying various candidate drugs, compounds, and treatment regimens to reduce the symptoms or effects of these diseases.

An additional embodiment of the invention relates to pharmaceutical compositions for treating disorders affected by expression or function of a novel IGERA isogene described herein. The pharmaceutical composition may comprise any of the following active ingredients: a polynucleotide comprising one of these novel IGERA isogenes; an antisense oligonucleotide directed against one of the novel IGERA isogenes, a polynucleotide encoding such an antisense oligonucleotide, or another compound which inhibits expression of a novel IGERA isogene described herein. Preferably, the

composition contains the active ingredient in a therapeutically effective amount. By therapeutically effective amount is meant that one or more of the symptoms relating to disorders affected by expression or function of a novel IGERA isogene is reduced and/or eliminated. The composition also comprises a pharmaceutically acceptable carrier, examples of which include, but are not limited to, saline, buffered saline, dextrose, and water. Those skilled in the art may employ a formulation most suitable for the active ingredient, whether it is a polynucleotide, oligonucleotide, protein, peptide or small molecule antagonist. The pharmaceutical composition may be administered alone or in combination with at least one other agent, such as a stabilizing compound. Administration of the pharmaceutical composition may be by any number of routes including, but not limited to oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, intradermal, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

For any composition, determination of the therapeutically effective dose of active ingredient and/or the appropriate route of administration is well within the capability of those skilled in the art. For example, the dose can be estimated initially either in cell culture assays or in animal models. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans. The exact dosage will be determined by the practitioner, in light of factors relating to the patient requiring treatment, including but not limited to severity of the disease state, general health, age, weight and gender of the patient, diet, time and frequency of administration, other drugs being taken by the patient, and tolerance/response to the treatment.

Information on the identity of genotypes and haplotypes for the IGERA gene of any particular individual as well as the frequency of such genotypes and haplotypes in any particular population of individuals is expected to be useful for a variety of basic research and clinical applications. Thus, the invention also provides compositions and methods for detecting the novel IGERA polymorphisms identified herein.

The compositions comprise at least one IGERA genotyping oligonucleotide. In one embodiment, a IGERA genotyping oligonucleotide is a probe or primer capable of hybridizing to a target region that is located close to, or that contains, one of the novel polymorphic sites described herein. As used herein, the term "oligonucleotide" refers to a polynucleotide molecule having less than about 100 nucleotides. A preferred oligonucleotide of the invention is 10 to 35 nucleotides long. More preferably, the oligonucleotide is between 15 and 30, and most preferably, between 20 and 25 nucleotides in length. The oligonucleotide may be comprised of any phosphorylation state of ribonucleotides, deoxyribonucleotides, and acyclic nucleotide derivatives, and other functionally equivalent derivatives. Alternatively, oligonucleotides may have a phosphate-free backbone, which may be comprised of linkages such as carboxymethyl, acetamidate, carbamate, polyamide (peptide

nucleic acid (PNA)) and the like (Varma, R. in Molecular Biology and Biotechnology, A Comprehensive Desk Reference, Ed. R. Meyers, VCH Publishers, Inc. (1995), pages 617-620). Oligonucleotides of the invention may be prepared by chemical synthesis using any suitable methodology known in the art, or may be derived from a biological sample, for example, by restriction digestion. The oligonucleotides may be labeled, according to any technique known in the art, including use of radiolabels, fluorescent labels, enzymatic labels, proteins, haptens, antibodies, sequence tags and the like.

Genotyping oligonucleotides of the invention must be capable of specifically hybridizing to a target region of a IGERA polynucleotide, i.e., a IGERA isogene. As used herein, specific hybridization means the oligonucleotide forms an anti-parallel double-stranded structure with the target region under certain hybridizing conditions, while failing to form such a structure when incubated with a non-target region or a non-IGERA polynucleotide under the same hybridizing conditions. Preferably, the oligonucleotide specifically hybridizes to the target region under conventional high stringency conditions. The skilled artisan can readily design and test oligonucleotide probes and primers suitable for detecting polymorphisms in the IGERA gene using the polymorphism information provided herein in conjunction with the known sequence information for the IGERA gene and routine techniques.

A nucleic acid molecule such as an oligonucleotide or polynucleotide is said to be a "perfect" or "complete" complement of another nucleic acid molecule if every nucleotide of one of the molecules is complementary to the nucleotide at the corresponding position of the other molecule. A nucleic acid molecule is "substantially complementary" to another molecule if it hybridizes to that molecule with sufficient stability to remain in a duplex form under conventional low-stringency conditions. Conventional hybridization conditions are described, for example, by Sambrook J. et al., in Molecular Cloning, A Laboratory Manual, 2<sup>nd</sup> Edition, Cold Spring Harbor Press, Cold Spring Harbor, NY (1989) and by Haymes, B.D. et al. in Nucleic Acid Hybridization, A Practical Approach, IRL Press, Washington, D.C. (1985). While perfectly complementary oligonucleotides are preferred for detecting polymorphisms, departures from complete complementarity are contemplated where such departures do not prevent the molecule from specifically hybridizing to the target region. For example, an oligonucleotide primer may have a non-complementary fragment at its 5' end, with the remainder of the primer being complementary to the target region. Alternatively, non-complementary nucleotides may be interspersed into the oligonucleotide probe or primer as long as the resulting probe or primer is still capable of specifically hybridizing to the target region.

Preferred genotyping oligonucleotides of the invention are allele-specific oligonucleotides. As used herein, the term allele-specific oligonucleotide (ASO) means an oligonucleotide that is able, under sufficiently stringent conditions, to hybridize specifically to one allele of a gene, or other locus, at a target region containing a polymorphic site while not hybridizing to the corresponding region in another allele(s). As understood by the skilled artisan, allele-specificity will depend upon a variety of readily optimized stringency conditions, including salt and formamide concentrations, as well as temperatures

for both the hybridization and washing steps. Examples of hybridization and washing conditions typically used for ASO probes are found in Kogan et al., "Genetic Prediction of Hemophilia A" in PCR Protocols, A Guide to Methods and Applications, Academic Press, 1990 and Ruano et al., 87 Proc. Natl. Acad. Sci. USA 6296-6300, 1990. Typically, an allele-specific oligonucleotide will be perfectly  
 5 complementary to one allele while containing a single mismatch for another allele.

Allele-specific oligonucleotide probes which usually provide good discrimination between different alleles are those in which a central position of the oligonucleotide probe aligns with the polymorphic site in the target region (e.g., approximately the 7<sup>th</sup> or 8<sup>th</sup> position in a 15 mer, the 8<sup>th</sup> or 9<sup>th</sup> position in a 16mer, the 10<sup>th</sup> or 11<sup>th</sup> position in a 20 mer). A preferred ASO probe for detecting IGERA  
 10 gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

	TGAAATATCAGATTT	(SEQ ID NO:4) and its complement,
	TGAAATAGCAGATTT	(SEQ ID NO:5) and its complement,
	ATTCTGCTCTCCCTT	(SEQ ID NO:6) and its complement,
15	ATTCTGCCCTCCCTT	(SEQ ID NO:7) and its complement,
	GATATGATACAGAAA	(SEQ ID NO:8) and its complement,
	GATATGACACAGAAA	(SEQ ID NO:9) and its complement,
	TACAGAAAACATTTT	(SEQ ID NO:10) and its complement,
	TACAGAATACATTTT	(SEQ ID NO:11) and its complement,
20	AATTACCCCTCCCAG	(SEQ ID NO:12) and its complement,
	AATTACCACTCCCAG	(SEQ ID NO:13) and its complement,
	ACTAATGTATCCTCT	(SEQ ID NO:14) and its complement,
	ACTAATGCATCCTCT	(SEQ ID NO:15) and its complement,
	GTATCCTCTCTGGAC	(SEQ ID NO:16) and its complement,
25	GTATCCTATCTGGAC	(SEQ ID NO:17) and its complement,
	TAATGAGCATGAATC	(SEQ ID NO:18) and its complement,
	TAATGAGTATGAATC	(SEQ ID NO:19) and its complement,
	AATCAAAACAGGGTC	(SEQ ID NO:20) and its complement,
	AATCAAAGCAGGGTC	(SEQ ID NO:21) and its complement,
30	AATGCCAAATTTGAA	(SEQ ID NO:22) and its complement,
	AATGCCAGATTTGAA	(SEQ ID NO:23) and its complement,
	AATGAGAGTGAACCT	(SEQ ID NO:24) and its complement,
	AATGAGAATGAACCT	(SEQ ID NO:25) and its complement,
	AGGCCTCTCATTTTTT	(SEQ ID NO:26) and its complement,
35	AGGCCTCCCATTTTTT	(SEQ ID NO:27) and its complement,
	TTTGGGAGGCTGAGG	(SEQ ID NO:28) and its complement,
	TTTGGGAAGCTGAGG	(SEQ ID NO:29) and its complement,

- 5      ACCATCCGGCTAACA (SEQ ID NO:30) and its complement,  
        ACCATCCAGCTAACA (SEQ ID NO:31) and its complement,  
        ATGCGTGGCTCTCTT (SEQ ID NO:32) and its complement,  
        ATGCGTGACTCTCTT (SEQ ID NO:33) and its complement,  
        TACTGTACGGGCAAA (SEQ ID NO:34) and its complement,  
        TACTGTATGGGCAAA (SEQ ID NO:35) and its complement,  
        AGCCTACCAGACTTG (SEQ ID NO:36) and its complement,  
        AGCCTACTAGACTTG (SEQ ID NO:37) and its complement,  
        ATGGTGATAGTAATA (SEQ ID NO:38) and its complement,  
 10     ATGGTGACAGTAATA (SEQ ID NO:39) and its complement,  
        TTCTGAACCCACATC (SEQ ID NO:40) and its complement,  
        TTCTGAAACCACATC (SEQ ID NO:41) and its complement,  
        CAATTGCTACTCAAT (SEQ ID NO:42) and its complement,  
        CAATTGCCACTCAAT (SEQ ID NO:43) and its complement,  
 15     AGCTTGCAATATACA (SEQ ID NO:44) and its complement,  
        AGCTTGCGATATACA (SEQ ID NO:45) and its complement,  
        TGAAACTGGTTAAGT (SEQ ID NO:46) and its complement, and  
        TGAAACTAGTTAAGT (SEQ ID NO:47) and its complement.

- 20      An allele-specific oligonucleotide primer of the invention has a 3' terminal nucleotide, or preferably a 3' penultimate nucleotide, that is complementary to only one nucleotide of a particular SNP, thereby acting as a primer for polymerase-mediated extension only if the allele containing that nucleotide is present. Allele-specific oligonucleotide primers hybridizing to either the coding or noncoding strand are contemplated by the invention. A preferred ASO primer for detecting IGERA  
 25     gene polymorphisms comprises a nucleotide sequence, listed 5' to 3', selected from the group consisting of:

- 30     AATAAATGAAATATC (SEQ ID NO:48); CTAAATAAATCTGAT (SEQ ID NO:49);  
        AATAAATGAAATAGC (SEQ ID NO:50); CTAAATAAATCTGCT (SEQ ID NO:51);  
        TGTTTTATTCTGCTC (SEQ ID NO:52); GGATGCAAGGGAGAG (SEQ ID NO:53);  
        TGTTTTATTCTGCCC (SEQ ID NO:54); GGATGCAAGGGAGGG (SEQ ID NO:55);  
        TAACCAGATATGATA (SEQ ID NO:56); AAATGTTTTCTGTAT (SEQ ID NO:57);  
        TAACCAGATATGACA (SEQ ID NO:58); AAATGTTTTCTGTGT (SEQ ID NO:59);  
        ATATGATACAGAAAA (SEQ ID NO:60); CAGAAGGAAATGTTT (SEQ ID NO:61);  
        ATATGATACAGAATA (SEQ ID NO:62); CAGAAGGAAATGTAT (SEQ ID NO:63);  
 35     AGATTCAATTACCCC (SEQ ID NO:64); GCCTCCCTGGGAGGG (SEQ ID NO:65);  
        AGATTCAATTACCAC (SEQ ID NO:66); GCCTCCCTGGGAGTG (SEQ ID NO:67);  
        CTGGACACTAATGTA (SEQ ID NO:68); GTCCAGAGAGGATAC (SEQ ID NO:69);

CTGGACACTAATGCA (SEQ ID NO:70); GTCCAGAGAGGATGC (SEQ ID NO:71);  
 ACTAATGTATCCTCT (SEQ ID NO:72); GCAAAAAGTCCAGAGA (SEQ ID NO:73);  
 ACTAATGTATCCTAT (SEQ ID NO:74); GCAAAAAGTCCAGATA (SEQ ID NO:75);  
 GCTTTCTAATGAGCA (SEQ ID NO:76); GGAACAGATTCATGC (SEQ ID NO:77);  
 5 GCTTTCTAATGAGTA (SEQ ID NO:78); GGAACAGATTCATAC (SEQ ID NO:79);  
 CCTAGAAATCAAAAC (SEQ ID NO:80); TGATAAGACCCTGTT (SEQ ID NO:81);  
 CCTAGAAATCAAAGC (SEQ ID NO:82); TGATAAGACCCTGCT (SEQ ID NO:83);  
 ATTGTGAATGCCAAA (SEQ ID NO:84); ACTGTCTTCAAATTT (SEQ ID NO:85);  
 ATTGTGAATGCCAGA (SEQ ID NO:86); ACTGTCTTCAAATCT (SEQ ID NO:87);  
 10 CAAGTTAATGAGAGT (SEQ ID NO:88); GTACACAGGTTCACT (SEQ ID NO:89);  
 CAAGTTAATGAGAAT (SEQ ID NO:90); GTACACAGGTTCAAT (SEQ ID NO:91);  
 GATTCAAGGCCTCTC (SEQ ID NO:92); GGTCTTAAAAATGAG (SEQ ID NO:93);  
 GATTCAAGGCCTCCC (SEQ ID NO:94); GGTCTTAAAAATGGG (SEQ ID NO:95);  
 CAGCACTTTGGGAGG (SEQ ID NO:96); CACCTGCCTCAGCCT (SEQ ID NO:97);  
 15 CAGCACTTTGGGAAG (SEQ ID NO:98); CACCTGCCTCAGCTT (SEQ ID NO:99);  
 ATCGAGACCATCCGG (SEQ ID NO:100); TCACCATGTTAGCCG (SEQ ID NO:101);  
 ATCGAGACCATCCAG (SEQ ID NO:102); TCACCATGTTAGCTG (SEQ ID NO:103);  
 TGCTCTATGCGTGGC (SEQ ID NO:104); AGAGAAAAGAGAGCC (SEQ ID NO:105);  
 TGCTCTATGCGTGAC (SEQ ID NO:106); AGAGAAAAGAGAGTC (SEQ ID NO:107);  
 20 ACCTACTACTGTACG (SEQ ID NO:108); CCACACTTTGCCCCGT (SEQ ID NO:109);  
 ACCTACTACTGTATG (SEQ ID NO:110); CCACACTTTGCCCCAT (SEQ ID NO:111);  
 CTGGAAAGCCTACCA (SEQ ID NO:112); TCATTGCAAGTCTGG (SEQ ID NO:113);  
 CTGGAAAGCCTACTA (SEQ ID NO:114); TCATTGCAAGTCTAG (SEQ ID NO:115);  
 TGTTAAATGGTGATA (SEQ ID NO:116); AGCAGGTATTACTAT (SEQ ID NO:117);  
 25 TGTTAAATGGTGACA (SEQ ID NO:118); AGCAGGTATTACTGT (SEQ ID NO:119);  
 TCAGACTTCTGAAAC (SEQ ID NO:120); GCTTAGGATGTGGGT (SEQ ID NO:121);  
 TCAGACTTCTGAAAC (SEQ ID NO:122); GCTTAGGATGTGGTT (SEQ ID NO:123);  
 CATCAGCAATTGCTA (SEQ ID NO:124); TTGACAATTGAGTAG (SEQ ID NO:125);  
 CATCAGCAATTGCCA (SEQ ID NO:126); TTGACAATTGAGTGG (SEQ ID NO:127);  
 30 AAACACAGCTTGCAA (SEQ ID NO:128); TTTCTATGTATATTG (SEQ ID NO:129);  
 AAACACAGCTTGCGA (SEQ ID NO:130); TTTCTATGTATATCG (SEQ ID NO:131);  
 ACTGAGTGAACTGG (SEQ ID NO:132); CATGCCACTTAACCA (SEQ ID NO:133);  
 ACTGAGTGAACTAG (SEQ ID NO:134); and CATGCCACTTAACCA (SEQ ID NO:135).

Other genotyping oligonucleotides of the invention hybridize to a target region located one to several  
 35 nucleotides downstream of one of the novel polymorphic sites identified herein. Such oligonucleotides  
 are useful in polymerase-mediated primer extension methods for detecting one of the novel  
 polymorphisms described herein and therefore such genotyping oligonucleotides are referred to herein



as "primer-extension oligonucleotides". In a preferred embodiment, the 3'-terminus of a primer-extension oligonucleotide is a deoxynucleotide complementary to the nucleotide located immediately adjacent to the polymorphic site. A particularly preferred oligonucleotide primer for detecting IGERA gene polymorphisms by primer extension terminates in a nucleotide sequence, listed 5' to 3', selected

5 from the group consisting of:

	AAATGAAATA (SEQ ID NO:136);	AATAAATCTG (SEQ ID NO:137);
	TTTATTCTGC (SEQ ID NO:138);	TGCAAGGGAG (SEQ ID NO:139);
	CCAGATATGA (SEQ ID NO:140);	TGTTTTCTGT (SEQ ID NO:141);
	TGATACAGAA (SEQ ID NO:142);	AAGGAAATGT (SEQ ID NO:143);
10	TTCAATTACC (SEQ ID NO:144);	TCCCTGGGAG (SEQ ID NO:145);
	GACACTAATG (SEQ ID NO:146);	CAGAGAGGAT (SEQ ID NO:147);
	AATGTATCCT (SEQ ID NO:148);	AAAGTCCAGA (SEQ ID NO:149);
	TTCTAATGAG (SEQ ID NO:150);	ACAGATTCAT (SEQ ID NO:151);
	AGAAATCAAA (SEQ ID NO:152);	TAAGACCCTG (SEQ ID NO:153);
15	GTGAATGCCA (SEQ ID NO:154);	GTCTTCAAAT (SEQ ID NO:155);
	GTTAATGAGA (SEQ ID NO:156);	CACAGGTTCA (SEQ ID NO:157);
	TCAAGGCCTC (SEQ ID NO:158);	CTTAAAAATG (SEQ ID NO:159);
	CACTTTGGGA (SEQ ID NO:160);	CTGCCTCAGC (SEQ ID NO:161);
	GAGACCATCC (SEQ ID NO:162);	CCATGTTAGC (SEQ ID NO:163);
20	TCTATGCGTG (SEQ ID NO:164);	GAAAAGAGAG (SEQ ID NO:165);
	TACTACTGTA (SEQ ID NO:166);	CACTTTGCCC (SEQ ID NO:167);
	GAAAGCCTAC (SEQ ID NO:168);	TTGCAAGTCT (SEQ ID NO:169);
	TAAATGGTGA (SEQ ID NO:170);	AGGTATTACT (SEQ ID NO:171);
	GACTTCTGAA (SEQ ID NO:172);	TAGGATGTGG (SEQ ID NO:173);
25	CAGCAATTGC (SEQ ID NO:174);	ACAATTGAGT (SEQ ID NO:175);
	CACAGCTTGC (SEQ ID NO:176);	CTATGTATAT (SEQ ID NO:177);
	GAGTGAAACT (SEQ ID NO:178);	and GCCACTTAAC (SEQ ID NO:179).

30 In some embodiments, a composition contains two or more differently labeled genotyping oligonucleotides for simultaneously probing the identity of nucleotides at two or more polymorphic sites. It is also contemplated that primer compositions may contain two or more sets of allele-specific primer pairs to allow simultaneous targeting and amplification of two or more regions containing a polymorphic site.

35 IGERA genotyping oligonucleotides of the invention may also be immobilized on or synthesized on a solid surface such as a microchip, bead, or glass slide (see, e.g., WO 98/20020 and WO 98/20019). Such immobilized genotyping oligonucleotides may be used in a variety of polymorphism detection assays, including but not limited to probe hybridization and polymerase

extension assays. Immobilized IGERA genotyping oligonucleotides of the invention may comprise an ordered array of oligonucleotides designed to rapidly screen a DNA sample for polymorphisms in multiple genes at the same time.

5 In another embodiment, the invention provides a kit comprising at least two genotyping oligonucleotides packaged in separate containers. The kit may also contain other components such as hybridization buffer (where the oligonucleotides are to be used as a probe) packaged in a separate container. Alternatively, where the oligonucleotides are to be used to amplify a target region, the kit may contain, packaged in separate containers, a polymerase and a reaction buffer optimized for primer extension mediated by the polymerase, such as PCR.

10 The above described oligonucleotide compositions and kits are useful in methods for genotyping and/or haplotyping the IGERA gene in an individual. As used herein, the terms "IGERA genotype" and "IGERA haplotype" mean the genotype or haplotype contains the nucleotide pair or nucleotide, respectively, that is present at one or more of the novel polymorphic sites described herein and may optionally also include the nucleotide pair or nucleotide present at one or more additional  
15 polymorphic sites in the IGERA gene. The additional polymorphic sites may be currently known polymorphic sites or sites that are subsequently discovered.

One embodiment of the genotyping method involves isolating from the individual a nucleic acid mixture comprising the two copies of the IGERA gene, or a fragment thereof, that are present in the individual, and determining the identity of the nucleotide pair at one or more of the polymorphic  
20 sites selected from PS1-22 in the two copies to assign a IGERA genotype to the individual. As will be readily understood by the skilled artisan, the two "copies" of a gene in an individual may be the same allele or may be different alleles. In a particularly preferred embodiment, the genotyping method comprises determining the identity of the nucleotide pair at each of PS1-22.

Typically, the nucleic acid mixture is isolated from a biological sample taken from the  
25 individual, such as a blood sample or tissue sample. Suitable tissue samples include whole blood, semen saliva, tears, urine, fecal material, sweat, buccal, skin and hair. The nucleic acid mixture may be comprised of genomic DNA, mRNA, or cDNA and, in the latter two cases, the biological sample must be obtained from an organ in which the IGERA gene is expressed. Furthermore it will be understood by the skilled artisan that mRNA or cDNA preparations would not be used to detect polymorphisms  
30 located in introns or in 5' and 3' nontranscribed regions. If a IGERA gene fragment is isolated, it must contain the polymorphic site(s) to be genotyped.

One embodiment of the haplotyping method comprises isolating from the individual a nucleic acid molecule containing only one of the two copies of the IGERA gene, or a fragment thereof, that is present in the individual and determining in that copy the identity of the nucleotide at one or more of  
35 the polymorphic sites PS1-22 in that copy to assign a IGERA haplotype to the individual. The nucleic acid may be isolated using any method capable of separating the two copies of the IGERA gene or fragment such as one of the methods described above for preparing IGERA isogenes, with targeted *in*

vivo cloning being the preferred approach. As will be readily appreciated by those skilled in the art, any individual clone will only provide haplotype information on one of the two IGERA gene copies present in an individual. If haplotype information is desired for the individual's other copy, additional IGERA clones will need to be examined. Typically, at least five clones should be examined to have more than a 90% probability of haplotyping both copies of the IGERA gene in an individual. In a particularly preferred embodiment, the nucleotide at each of PS1-22 is identified.

In a preferred embodiment, a IGERA haplotype pair is determined for an individual by identifying the phased sequence of nucleotides at one or more of the polymorphic sites selected from PS1-22 in each copy of the IGERA gene that is present in the individual. In a particularly preferred embodiment, the haplotyping method comprises identifying the phased sequence of nucleotides at each of PS1-22 in each copy of the IGERA gene. When haplotyping both copies of the gene, the identifying step is preferably performed with each copy of the gene being placed in separate containers. However, it is also envisioned that if the two copies are labeled with different tags, or are otherwise separately distinguishable or identifiable, it could be possible in some cases to perform the method in the same container. For example, if first and second copies of the gene are labeled with different first and second fluorescent dyes, respectively, and an allele-specific oligonucleotide labeled with yet a third different fluorescent dye is used to assay the polymorphic site(s), then detecting a combination of the first and third dyes would identify the polymorphism in the first gene copy while detecting a combination of the second and third dyes would identify the polymorphism in the second gene copy.

In both the genotyping and haplotyping methods, the identity of a nucleotide (or nucleotide pair) at a polymorphic site(s) may be determined by amplifying a target region(s) containing the polymorphic site(s) directly from one or both copies of the IGERA gene, or fragment thereof, and the sequence of the amplified region(s) determined by conventional methods. It will be readily appreciated by the skilled artisan that only one nucleotide will be detected at a polymorphic site in individuals who are homozygous at that site, while two different nucleotides will be detected if the individual is heterozygous for that site. The polymorphism may be identified directly, known as positive-type identification, or by inference, referred to as negative-type identification. For example, where a SNP is known to be guanine and cytosine in a reference population, a site may be positively determined to be either guanine or cytosine for an individual homozygous at that site, or both guanine and cytosine, if the individual is heterozygous at that site. Alternatively, the site may be negatively determined to be not guanine (and thus cytosine/cytosine) or not cytosine (and thus guanine/guanine).

In addition, the identity of the allele(s) present at any of the novel polymorphic sites described herein may be indirectly determined by genotyping a polymorphic site not disclosed herein that is in linkage disequilibrium with the polymorphic site that is of interest. Two sites are said to be in linkage disequilibrium if the presence of a particular variant at one site enhances the predictability of another variant at the second site (Stevens, JC 1999, *Mol. Diag.* 4: 309-17). Polymorphic sites in linkage disequilibrium with the presently disclosed polymorphic sites may be located in regions of the gene or

in other genomic regions not examined herein. Genotyping of a polymorphic site in linkage disequilibrium with the novel polymorphic sites described herein may be performed by, but is not limited to, any of the above-mentioned methods for detecting the identity of the allele at a polymorphic site.

5           The target region(s) may be amplified using any oligonucleotide-directed amplification method, including but not limited to polymerase chain reaction (PCR) (U.S. Patent No. 4,965,188), ligase chain reaction (LCR) (Barany et al., *Proc. Natl. Acad. Sci. USA* 88:189-193, 1991; WO90/01069), and oligonucleotide ligation assay (OLA) (Landegren et al., *Science* 241:1077-1080, 1988).

10           Oligonucleotides useful as primers or probes in such methods should specifically hybridize to a region of the nucleic acid that contains or is adjacent to the polymorphic site. Typically, the oligonucleotides are between 10 and 35 nucleotides in length and preferably, between 15 and 30 nucleotides in length. Most preferably, the oligonucleotides are 20 to 25 nucleotides long. The exact length of the oligonucleotide will depend on many factors that are routinely considered and practiced by the skilled artisan.

15           Other known nucleic acid amplification procedures may be used to amplify the target region including transcription-based amplification systems (U.S. Patent No. 5,130,238; EP 329,822; U.S. Patent No. 5,169,766, WO89/06700) and isothermal methods (Walker et al., *Proc. Natl. Acad. Sci. USA* 89:392-396, 1992).

20           A polymorphism in the target region may also be assayed before or after amplification using one of several hybridization-based methods known in the art. Typically, allele-specific oligonucleotides are utilized in performing such methods. The allele-specific oligonucleotides may be used as differently labeled probe pairs, with one member of the pair showing a perfect match to one variant of a target sequence and the other member showing a perfect match to a different variant. In some embodiments, more than one polymorphic site may be detected at once using a set of allele-specific oligonucleotides or oligonucleotide pairs. Preferably, the members of the set have melting  
25           temperatures within 5°C, and more preferably within 2°C, of each other when hybridizing to each of the polymorphic sites being detected.

30           Hybridization of an allele-specific oligonucleotide to a target polynucleotide may be performed with both entities in solution, or such hybridization may be performed when either the oligonucleotide or the target polynucleotide is covalently or noncovalently affixed to a solid support. Attachment may be mediated, for example, by antibody-antigen interactions, poly-L-Lys, streptavidin or avidin-biotin, salt bridges, hydrophobic interactions, chemical linkages, UV cross-linking baking, etc. Allele-specific oligonucleotides may be synthesized directly on the solid support or attached to the solid support subsequent to synthesis. Solid-supports suitable for use in detection methods of the invention include  
35           substrates made of silicon, glass, plastic, paper and the like, which may be formed, for example, into wells (as in 96-well plates), slides, sheets, membranes, fibers, chips, dishes, and beads. The solid support may be treated, coated or derivatized to facilitate the immobilization of the allele-specific

oligonucleotide or target nucleic acid.

The genotype or haplotype for the IGERA gene of an individual may also be determined by hybridization of a nucleic sample containing one or both copies of the gene to nucleic acid arrays and subarrays such as described in WO 95/11995. The arrays would contain a battery of allele-specific  
5 oligonucleotides representing each of the polymorphic sites to be included in the genotype or haplotype.

The identity of polymorphisms may also be determined using a mismatch detection technique, including but not limited to the RNase protection method using riboprobes (Winter et al., Proc. Natl. Acad. Sci. USA 82:7575, 1985; Meyers et al., Science 230:1242, 1985) and proteins which recognize nucleotide mismatches, such as the E. coli mutS protein (Modrich, P. Ann. Rev. Genet. 25:229-253,  
10 1991). Alternatively, variant alleles can be identified by single strand conformation polymorphism (SSCP) analysis (Orita et al., Genomics 5:874-879, 1989; Humphries et al., in Molecular Diagnosis of Genetic Diseases, R. Elles, ed., pp. 321-340, 1996) or denaturing gradient gel electrophoresis (DGGE) (Wartell et al., Nucl. Acids Res. 18:2699-2706, 1990; Sheffield et al., Proc. Natl. Acad. Sci. USA 86:232-236, 1989).

A polymerase-mediated primer extension method may also be used to identify the  
15 polymorphism(s). Several such methods have been described in the patent and scientific literature and include the "Genetic Bit Analysis" method (WO92/15712) and the ligase/polymerase mediated genetic bit analysis (U.S. Patent 5,679,524. Related methods are disclosed in WO91/02087, WO90/09455, WO95/17676, U.S. Patent Nos. 5,302,509, and 5,945,283. Extended primers containing a  
20 polymorphism may be detected by mass spectrometry as described in U.S. Patent No. 5,605,798. Another primer extension method is allele-specific PCR (Ruano et al., Nucl. Acids Res. 17:8392, 1989; Ruano et al., Nucl. Acids Res. 19, 6877-6882, 1991; WO 93/22456; Turki et al., J. Clin. Invest. 95:1635-1641, 1995). In addition, multiple polymorphic sites may be investigated by simultaneously amplifying multiple regions of the nucleic acid using sets of allele-specific primers as described in  
25 Wallace et al. (WO89/10414).

In another aspect of the invention, an individual's IGERA haplotype pair is predicted from its IGERA genotype using information on haplotype pairs known to exist in a reference population. In its broadest embodiment, the haplotyping prediction method comprises identifying a IGERA genotype for the individual at two or more polymorphic sites selected from PS1-22, enumerating all possible  
30 haplotype pairs which are consistent with the genotype, accessing data containing IGERA haplotype pairs identified in a reference population, and assigning a haplotype pair to the individual that is consistent with the data. In one embodiment, the reference haplotype pairs include the IGERA haplotype pairs shown in Table 4.

Generally, the reference population should be composed of randomly-selected individuals  
35 representing the major ethnogeographic groups of the world. A preferred reference population for use in the methods of the present invention comprises an approximately equal number of individuals from Caucasian, African American, Asian and Hispanic-Latino population groups with the minimum number

of each group being chosen based on how rare a haplotype one wants to be guaranteed to see. For example, if one wants to have a  $q\%$  chance of not missing a haplotype that exists in the population at a  $p\%$  frequency of occurring in the reference population, the number of individuals ( $n$ ) who must be sampled is given by  $2n = \log(1-q)/\log(1-p)$  where  $p$  and  $q$  are expressed as fractions. A preferred reference population allows the detection of any haplotype whose frequency is at least 10% with about 99% certainty and comprises about 20 unrelated individuals from each of the four population groups named above. A particularly preferred reference population includes a 3-generation family representing one or more of the four population groups to serve as controls for checking quality of haplotyping procedures.

In a preferred embodiment, the haplotype frequency data for each ethnogeographic group is examined to determine whether it is consistent with Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (D.L. Hartl et al., Principles of Population Genomics, Sinauer Associates (Sunderland, MA), 3<sup>rd</sup> Ed., 1997) postulates that the frequency of finding the haplotype pair  $H_1 / H_2$  is equal to  $p_{H-W}(H_1 / H_2) = 2p(H_1)p(H_2)$  if  $H_1 \neq H_2$  and  $p_{H-W}(H_1 / H_2) = p(H_1)p(H_2)$  if  $H_1 = H_2$ . A statistically significant difference between the observed and expected haplotype frequencies could be due to one or more factors including significant inbreeding in the population group, strong selective pressure on the gene, sampling bias, and/or errors in the genotyping process. If large deviations from Hardy-Weinberg equilibrium are observed in an ethnogeographic group, the number of individuals in that group can be increased to see if the deviation is due to a sampling bias. If a larger sample size does not reduce the difference between observed and expected haplotype pair frequencies, then one may wish to consider haplotyping the individual using a direct haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

In one embodiment of this method for predicting a IGERA haplotype pair, the assigning step involves performing the following analysis. First, each of the possible haplotype pairs is compared to the haplotype pairs in the reference population. Generally, only one of the haplotype pairs in the reference population matches a possible haplotype pair and that pair is assigned to the individual. Occasionally, only one haplotype represented in the reference haplotype pairs is consistent with a possible haplotype pair for an individual, and in such cases the individual is assigned a haplotype pair containing this known haplotype and a new haplotype derived by subtracting the known haplotype from the possible haplotype pair. In rare cases, either no haplotypes in the reference population are consistent with the possible haplotype pairs, or alternatively, multiple reference haplotype pairs are consistent with the possible haplotype pairs. In such cases, the individual is preferably haplotyped using a direct molecular haplotyping method such as, for example, CLASPER System™ technology (U.S. Patent No. 5,866,404), SMD, or allele-specific long-range PCR (Michalotos-Beloin et al., Nucleic Acids Res. 24:4841-4843, 1996).

The invention also provides a method for determining the frequency of a IGERA genotype or IGERA haplotype in a population. The method comprises determining the genotype or the haplotype pair for the IGERA gene that is present in each member of the population, wherein the genotype or haplotype comprises the nucleotide pair or nucleotide detected at one or more of the polymorphic sites PS1-22 in the IGERA gene; and calculating the frequency any particular genotype or haplotype is found in the population. The population may be a reference population, a family population, a same sex population, a population group, a trait population (e.g., a group of individuals exhibiting a trait of interest such as a medical condition or response to a therapeutic treatment).

In another aspect of the invention, frequency data for IGERA genotypes and/or haplotypes found in a reference population are used in a method for identifying an association between a trait and a IGERA genotype or a IGERA haplotype. The trait may be any detectable phenotype, including but not limited to susceptibility to a disease or response to a treatment. The method involves obtaining data on the frequency of the genotype(s) or haplotype(s) of interest in a reference population as well as in a population exhibiting the trait. Frequency data for one or both of the reference and trait populations may be obtained by genotyping or haplotyping each individual in the populations using one of the methods described above. The haplotypes for the trait population may be determined directly or, alternatively, by the predictive genotype to haplotype approach described above. In another embodiment, the frequency data for the reference and/or trait populations is obtained by accessing previously determined frequency data, which may be in written or electronic form. For example, the frequency data may be present in a database that is accessible by a computer. Once the frequency data is obtained, the frequencies of the genotype(s) or haplotype(s) of interest in the reference and trait populations are compared. In a preferred embodiment, the frequencies of all genotypes and/or haplotypes observed in the populations are compared. If a particular genotype or haplotype for the IGERA gene is more frequent in the trait population than in the reference population at a statistically significant amount, then the trait is predicted to be associated with that IGERA genotype or haplotype. Preferably, the IGERA genotype or haplotype being compared in the trait and reference populations is selected from the full-genotypes and full-haplotypes shown in Tables 4 and 5, respectively, or from sub-genotypes and sub-haplotypes derived from these genotypes and haplotypes.

In a preferred embodiment of the method, the trait of interest is a clinical response exhibited by a patient to some therapeutic treatment, for example, response to a drug targeting IGERA or response to a therapeutic treatment for a medical condition. As used herein, "medical condition" includes but is not limited to any condition or disease manifested as one or more physical and/or psychological symptoms for which treatment is desirable, and includes previously and newly identified diseases and other disorders. As used herein the term "clinical response" means any or all of the following: a quantitative measure of the response, no response, and adverse response (i.e., side effects).

In order to deduce a correlation between clinical response to a treatment and a IGERA genotype or haplotype, it is necessary to obtain data on the clinical responses exhibited by a population of

individuals who received the treatment, hereinafter the "clinical population". This clinical data may be obtained by analyzing the results of a clinical trial that has already been run and/or the clinical data may be obtained by designing and carrying out one or more new clinical trials. As used herein, the term "clinical trial" means any research study designed to collect clinical data on responses to a particular treatment, and includes but is not limited to phase I, phase II and phase III clinical trials. Standard methods are used to define the patient population and to enroll subjects.

It is preferred that the individuals included in the clinical population have been graded for the existence of the medical condition of interest. This is important in cases where the symptom(s) being presented by the patients can be caused by more than one underlying condition, and where treatment of the underlying conditions are not the same. An example of this would be where patients experience breathing difficulties that are due to either asthma or respiratory infections. If both sets were treated with an asthma medication, there would be a spurious group of apparent non-responders that did not actually have asthma. These people would affect the ability to detect any correlation between haplotype and treatment outcome. This grading of potential patients could employ a standard physical exam or one or more lab tests. Alternatively, grading of patients could use haplotyping for situations where there is a strong correlation between haplotype pair and disease susceptibility or severity.

The therapeutic treatment of interest is administered to each individual in the trial population and each individual's response to the treatment is measured using one or more predetermined criteria. It is contemplated that in many cases, the trial population will exhibit a range of responses and that the investigator will choose the number of responder groups (e.g., low, medium, high) made up by the various responses. In addition, the IGERA gene for each individual in the trial population is genotyped and/or haplotyped, which may be done before or after administering the treatment.

After both the clinical and polymorphism data have been obtained, correlations between individual response and IGERA genotype or haplotype content are created. Correlations may be produced in several ways. In one method, individuals are grouped by their IGERA genotype or haplotype (or haplotype pair) (also referred to as a polymorphism group), and then the averages and standard deviations of clinical responses exhibited by the members of each polymorphism group are calculated.

These results are then analyzed to determine if any observed variation in clinical response between polymorphism groups is statistically significant. Statistical analysis methods which may be used are described in L.D. Fisher and G. vanBelle, "Biostatistics: A Methodology for the Health Sciences", Wiley-Interscience (New York) 1993. This analysis may also include a regression calculation of which polymorphic sites in the PTGS2 gene give the most significant contribution to the differences in phenotype. One regression model useful in the invention is described in the PCT Application entitled "Methods for Obtaining and Using Haplotype Data", filed June 26, 2000.

Correlations may also be analyzed using analysis of variation (ANOVA) techniques to determine how much of the variation in the clinical data is explained by different subsets of the



polymorphic sites in the IGERA gene. As described in PCT Application entitled "Methods for Obtaining and Using Haplotype Data", filed June 26, 2000, ANOVA is used to test hypotheses about whether a response variable is caused by or correlated with one or more traits or variables that can be measured (Fisher and vanBelle, supra, Ch. 10).

5 From the analyses described above, a mathematical model may be readily constructed by the skilled artisan that predicts clinical response as a function of IGERA genotype or haplotype content. Preferably, the model is validated in one or more follow-up clinical trials designed to test the model.

The identification of an association between a clinical response and a genotype or haplotype (or haplotype pair) for the IGERA gene may be the basis for designing a diagnostic method to determine  
10 those individuals who will or will not respond to the treatment, or alternatively, will respond at a lower level and thus may require more treatment, i.e., a greater dose of a drug. The diagnostic method may take one of several forms: for example, a direct DNA test (i.e., genotyping or haplotyping one or more of the polymorphic sites in the IGERA gene), a serological test, or a physical exam measurement. The only requirement is that there be a good correlation between the diagnostic test results and the  
15 underlying IGERA genotype or haplotype that is in turn correlated with the clinical response. In a preferred embodiment, this diagnostic method uses the predictive haplotyping method described above.

Any or all analytical and mathematical operations involved in practicing the methods of the present invention may be implemented by a computer. In addition, the computer may execute a program that generates views (or screens) displayed on a display device and with which the user can  
20 interact to view and analyze large amounts of information relating to the IGERA gene and its genomic variation, including chromosome location, gene structure, and gene family, gene expression data, polymorphism data, genetic sequence data, and clinical data population data (e.g., data on ethnogeographic origin, clinical responses, genotypes, and haplotypes for one or more populations). The IGERA polymorphism data described herein may be stored as part of a relational database (e.g., an  
25 instance of an Oracle database or a set of ASCII flat files). These polymorphism data may be stored on the computer's hard drive or may, for example, be stored on a CD ROM or on one or more other storage devices accessible by the computer. For example, the data may be stored on one or more databases in communication with the computer via a network.

Preferred embodiments of the invention are described in the following examples. Other  
30 embodiments within the scope of the claims herein will be apparent to one skilled in the art from consideration of the specification or practice of the invention as disclosed herein. It is intended that the specification, together with the examples, be considered exemplary only, with the scope and spirit of the invention being indicated by the claims which follow the examples.

35

#### EXAMPLES

The Examples herein are meant to exemplify the various aspects of carrying out the invention and are not intended to limit the scope of the invention in any way. The Examples do not include

detailed descriptions for conventional methods employed, such as in the performance of genomic DNA isolation, PCR and sequencing procedures. Such methods are well-known to those skilled in the art and are described in numerous publications, for example, Sambrook, Fritsch, and Maniatis, "Molecular Cloning: A Laboratory Manual", 2<sup>nd</sup> Edition, Cold Spring Harbor Laboratory Press, USA, (1989).

5

### Example 1

This example illustrates examination of various regions of the IGERA gene for polymorphic sites.

#### 10 Amplification of Target Regions

The following target regions of the IGERA gene were amplified using the PCR primer pairs listed below, with the sequences presented in the 5' to 3' direction and nucleotide positions shown for each region corresponding to the indicated GenBank Accession No.

Accession Number: L14075

#### 15 Fragment 1

Forward Primer

605-627 AAGAAAAGCGTTGGTAGCTCTGG (SEQ ID NO:180)

Reverse Primer

Complement of 1424-1401 CACCCACAGTAAAGGTTCTACCC (SEQ ID NO:181)

20 PCR product 820 nt

#### Fragment 2

Forward Primer

1033-1055 ATGCCTCTCTCTCACCAGATTCC (SEQ ID NO:182)

25 Reverse Primer

Complement of 1507-1485 CTTCCTCTGCTTTCTAGCTTGG (SEQ ID NO:183)

PCR product 475 nt

#### Fragment 3

30 Forward Primer

1074-1096 GGGATAGGGAGTGGAGTAAGTGG (SEQ ID NO:184)

Reverse Primer

Complement of 1617-1592 TCCTCTACCCTCATTACCTTGGTAGG (SEQ ID NO:185)

PCR product 544 nt

35

#### Fragment 4

Forward Primer

1605-1628 TGAGGGTAGAGGAGAGAAAGAAGC (SEQ ID NO:186)

Reverse Primer

40 Complement of 1995-1970 GAGGAGAGAATGACTTGAGAGAATGC (SEQ ID NO:187)

PCR product 391 nt

#### Fragment 5

Forward Primer

45 2637-2658 CCTGTCTTTCTCCCTGTGTTGG (SEQ ID NO:188)

Reverse Primer

Complement of 3205-3183 CACTCTGGTGTCTAACCCTTGG (SEQ ID NO:189)

PCR product 569 nt

## Fragment 6

## Forward Primer

2839-2862 TCTTCTTGAAGTCCCTCAGAAACC (SEQ ID NO:190)

## Reverse Primer

- 5 Complement of 3353-3331 AGATGGGGTTTCACCATGTTAGC (SEQ ID NO:191)  
 PCR product 515 nt

## Fragment 7

## Forward Primer

- 10 4645-4668 TGTTCATGTATGGACTCATCAGG (SEQ ID NO:192)

## Reverse Primer

Complement of 5218-5196 CTCTCTTCCTTCCCCTGCTATGG (SEQ ID NO:193)  
 PCR product 574 nt

- 15 Fragment 8

## Forward Primer

4813-4834 GTTCTGACACATGCTCTATGC (SEQ ID NO:194)

## Reverse Primer

- 20 Complement of 5463-5443 TCTGTTATGCTTGGGTAGTGC (SEQ ID NO:195)  
 PCR product 651 nt

## Fragment 9

## Forward Primer

6486-6507 GCACCAACAGAGCAACTCAACC (SEQ ID NO:196)

- 25 Reverse Primer

Complement of 7212-7187 CCAATCTAGAACTTCATGGTCCTTGC (SEQ ID NO:197)  
 PCR product 727 nt

## Fragment 10

## Forward Primer

- 30 6709-6730 TGTTGGTGGTGATTCTGTTTGC (SEQ ID NO:198)

## Reverse Primer

Complement of 7359-7336 TCTTGAGACTGTCCCTGATTCTGC (SEQ ID NO:199)  
 PCR product 651 nt

35

These primer pairs were used in PCR reactions containing genomic DNA isolated from immortalized cell lines for each member of the Index Repository. The PCR reactions were carried out under the following conditions:

	Reaction volume	= 20 $\mu$ l
40	10 x Advantage 2 Polymerase reaction buffer (Clontech)	= 2 $\mu$ l
	100 ng of human genomic DNA	= 1 $\mu$ l
	10 mM dNTP	= 0.4 $\mu$ l
	Advantage 2 Polymerase enzyme mix (Clontech)	= 0.2 $\mu$ l
	Forward Primer (10 $\mu$ M)	= 0.4 $\mu$ l
45	Reverse Primer (10 $\mu$ M)	= 0.4 $\mu$ l
	Water	=15.6 $\mu$ l

## Amplification profile:

50	94°C - 2 min.	1 cycle
	94°C - 30 sec.	} 10 cycles
	70°C - 45 sec.	
	72°C - 1 min.	

94°C - 30 sec.  
64°C - 45 sec.  
72°C - 1 min. } 35 cycles

### Sequencing of PCR Products

5 The PCR products were purified by Solid Phase Reversible Immobilization using the protocol developed by the Whitehead Genome Center. A detailed protocol can be found at [http://www.genome.wi.mit.edu/sequencing/protocols/pure/SPRI\\_protocol.html](http://www.genome.wi.mit.edu/sequencing/protocols/pure/SPRI_protocol.html).

Briefly, five µl of carboxyl coated magnetic beads (10 mg/ml) and 60 µl of HYB BUFFER (2.5M NaCl/20% PEG 8000) were added to each PCR reaction mixture (20 µl). The reaction mixture  
10 was mixed well and incubated at room temperature (RT) for 10 min. The microtitre plate was placed on a magnet for 2 min and the beads washed twice with 150 µl of 70% EtOH. The beads were air dried for 2 min and the DNA was eluted in 25 µl of distilled water and incubated at RT for 5 min. The beads were magnetically separated and the supernatant removed for testing and sequencing.

The purified PCR products were sequenced in both directions using the primer sets described  
15 previously or those listed, in the 5' to 3' direction, below.

Accession Number: L14075

#### Fragment 1

##### Forward Primer

756-775 AGTTGGCACCCCAAAACAAG (SEQ ID NO:200)

##### Reverse Primer

20 Complement of 1299-1280 TGGCAGGAGCCATCTTCTTC (SEQ ID NO:201)

#### Fragment 2

##### Forward Primer

25 1068-1087 GGAGGTGGGATAGGGAGTGG (SEQ ID NO:202)

##### Reverse Primer

Complement of 1471-1452 TCCCTGGGAAATGCCCAATA (SEQ ID NO:203)

#### Fragment 3

##### Forward Primer

30 1114-1133 CAGTTGGGCACCATCCTGAA (SEQ ID NO:204)

##### Reverse Primer

Complement of 1581-1561 TCTGGAAGATGCCAGAGCAAA (SEQ ID NO:205)

#### Fragment 4

##### Forward Primer

35 1652-1673 CCTGAAAAGACGGTTGGTCCTT (SEQ ID NO:206)

##### Reverse Primer

Complement of 1942-1923 AGGCAAGGTGGAGAGGGAAA (SEQ ID NO:207)

40

#### Fragment 5

##### Forward Primer

2661-2680 GTTCCCTGGGGCACCAATAC (SEQ ID NO:208)

##### Reverse Primer

45 Complement of 3160-3140 TCAGATGAGCCATCCCTCACA (SEQ ID NO:209)

## Fragment 6

Forward Primer

2871-2892 CCTTGAACCCTCCATGGAATAG (SEQ ID NO:210)

Reverse Primer

5 Complement of 3246-3227 CAGCCAATGCAGGGGTCTTA (SEQ ID NO:211)

## Fragment 7

Forward Primer

4685-4704 TGTGGCCCCAGACTGACTTT (SEQ ID NO:212)

10 Reverse Primer

Complement of 5152-5133 TGTTGAGGGGCTCAGACTCA (SEQ ID NO:213)

## Fragment 8

Forward Primer

15 4930-4950 TCTGCTGAGGTGGTGATGGAG (SEQ ID NO:214)

Reverse Primer

Complement of 5311-5292 CACCCAGGTCTCCTCATTGC (SEQ ID NO:215)

## Fragment 9

20 Forward Primer

6541-6559 GGATGCCACATCACGCTAA (SEQ ID NO:216)

Reverse Primer

Complement of 7013-6992 CATGCCACTTAACCAGTTTCAC (SEQ ID NO:217)

25 Fragment 10

Forward Primer

6791-6812 GAGAACCAGGAAAGGCTTCAGA (SEQ ID NO:218)

Reverse Primer

30 Complement of 7284-7265 TCCACCTCACTGGCATCCTC (SEQ NO:219)

Analysis of Sequences for Polymorphic Sites

Sequences were analyzed for the presence of polymorphisms using the Polyphred program  
 35 (Nickerson et al., Nucleic Acids Res. 14:2745-2751, 1997). The presence of a polymorphism was confirmed on both strands. The polymorphisms and their locations in the IGERA gene are listed in Table 3 below.

Table 3. Polymorphic Sites Identified in the IGERA Gene

	Polymorphic Site Number	Nucleotide Position	Reference Allele	Variant Allele
	PS1	872(Acc#L14075)	T	G
5	PS2	943(Acc#L14075)	T	C
	PS3	1192(Acc#L14075)	T	C
	PS4	1199(Acc#L14075)	A	T
	PS5	1363(Acc#L14075)	C	A
	PS6	1754(Acc#L14075)	T	C
10	PS7	1760(Acc#L14075)	C	A
	PS8	1896(Acc#L14075)	C	T
	PS9	2708(Acc#L14075)	A	G
	PS10	3024(Acc#L14075)	A	G
	PS11	3075(Acc#L14075)	G	A
15	PS12	3220(Acc#L14075)	T	C
	PS13	3286(Acc#L14075)	G	A
	PS14	3330(Acc#L14075)	G	A
	PS15	4838(Acc#L14075)	G	A
	PS16	5108(Acc#L14075)	C	T
20	PS17	5285(Acc#L14075)	C	T
	PS18	5363(Acc#L14075)	T	C
	PS19	6821(Acc#L14075)	C	A
	PS20	6911(Acc#L14075)	T	C
	PS21	6936(Acc#L14075)	A	G
25	PS22	7000(Acc#L14075)	G	A

## Example 2

This example illustrates analysis of the IGERA polymorphisms identified in the Index Repository for human genotypes and haplotypes.

30 The different genotypes containing these polymorphisms that were observed in the reference population are shown in Table 4 below, with the haplotype pair indicating the combination of haplotypes determined for the individual using the haplotype derivation protocol described below. In Table 4, homozygous positions are indicated by one nucleotide and heterozygous positions are indicated by two nucleotides. Missing nucleotides in any given genotype in Table 4 can typically be  
 35 inferred based on linkage disequilibrium and/or Mendelian inheritance.

Table 4. Genotypes and Haplotype Pairs Observed in the IGERA Gene																								
Genotype Number		Polymorphic Sites																				HAP PAIR		
		PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20			PS21
1	T	C	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T	T	C	T	A	G	1
2	T	C/T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T/C	T	C	T	A	G	2
3	T	C/T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T/C	T	C	T	A	G/A	3
4	T	C	T/C	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T	T	C	T	A	G	4
5	T/G	C	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T	T	C	T	A	G	5
6	T	C	T	A	C	T	C	C/T	A	A/G	G	T	G	G	G/A	C	C	T	T/C	C	T	A/G	G	7
7	T/G	C	T	A	C	T/C	C	C	A	A	G	T/C	G	G	G	C	C	T	T	C	T/C	A	G	11
8	T	C	T	A	C	T	C	C	A	A	G	T	G	G	-	-	T	T	C/A	T	A	G	12	
9	T	C	T	A	C	T	C/A	C/T	A	A	G	T	G	G	G	C	C	T	T	C	T	A	G	15
10	T	C	T	A	C	T	C	C	A	A	G	T	G	G/A	G	C	C	T	T	C/A	T	A	G	16
11	T	C	T	A	C	T	C	C/T	A	A/G	G	T	G	G	G	C	C	T	T/C	C	T	A	G	17
12	T	C/T	T	A	C/A	T	C	C/T	A	A	G	T	G	G	G/A	C	C	T/C	T/C	C	T	A	G	20
13	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	C	T	C	T	A	G	22
14	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	C	T	C	T	A	G/A	23
15	T	C/T	T/C	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T/C	T	C	T	A	G	24
16	T	C/T	T	A	C	T	C	C/T	A	A	G	T	G	G	G/A	C	C	T/C	T/C	C	T	A	G	26
17	T	T	T	A	C	T	C	C	A/G	A	G	T	G	G	G	C	C	C	T	-	T	A	G	29
18	T	C/T	T	A	C	T	C	C/T	A	A	G/A	T	G/A	G	G	C	C	T/C	T	C	T	A	G	210
19	T	C/T	T	A	C	T	C	C/T	-	A	G	T	G	G	G/A	C	C	T/C	T/C	C	T	A/G	G	213
20	T	T	T	A	C	T	C	C/T	A	A	G	T	G	G	G	C	C	T/C	T	C	T	A	G/A	214
21	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T	T	C	T	A	A	33
22	T	C/T	T/C	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T/C	T	C	T	A	G/A	34

Table 4 (cont'd). Genotypes and Haplotype Pairs Observed in the IGRA Gene																							
Genotype Number	Polymorphic Sites																						
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20	PS21	PS22	
23	T/G	C/T	T	A	C	T	C	C	A	A	G	-	-	-	G	C	T/C	T	C	T	A	G/A	3
24	T	C/T	T	A	C	T	C	-	A	A	G	T	G	G	G/A	C	T/C	T/C	C	T	A	G/A	3
25	T	C/T	T	A	C	T	C	C/T	A	A	G	T	G	G	G/A	C	T/C	T/C	C	T	A	G/A	3
26	T	T	T	A	C	T	C	C	A/G	A	G	-	-	-	G	C	C	T	C	T	A	G/A	3
27	T	C/T	T	A	C	T	C	-	A	A	G	T	G	G	G	C	T/C	T	C/A	T	A	G/A	3
28	T	C/T	T	A	C	T	C/A	C/T	A	A	G	T	G	G	G	C	T/C	T	C	T	A	G/A	3
29	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	C/T	C	T	C	T	A	A	3
30	T	C	C	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	C	T	A	G	4
31	T/G	C	T/C	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	C	T	A	G	4
32	T	C	T/C	A	C	T	C	C/T	A	A	G	T	G	G	G/A	C	T	T/C	C	T	A	G	4
33	T	C	T/C	A	C	T	C	C/T	A	A	G	T	G	G	G	C	T	T	C	T	A	G	4
34	T/G	C	T/C	A	C	T/C	C	C	A	A	G	T/C	G	G	G	C	T	T	C	T/C	A	G	4
35	T	C	T/C	A	C	T	C	C/T	A	A	G	T	G	G	G/A	C	-	-	C	T	A/G	G	4
36	G	C	T	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	C	T	A	G	5
37	G	C	T	A	C	T/C	C	C	A	A	G	T/C	G	G	G	C	T	T	C	T/C	A	G	5
38	T/G	C	T	A	C	T	C/A	C/T	A	A	G	T	G	G	G	C	T	T	C	T	A	G	5
39	T	C	T	A	C	T	C	-	A	A	G	T	G	G	A	C	T	C	C	T	A	G	6
40	T	C	T	A	C	T	C	T	A	A/G	G	T	G	G	A	C	T	C	C	T	A/G	G	6
41	T	C	T	A	C	T	C	T	A	A	G	T	G	G	G/A	C	T	T/C	C	T	A	G	6
42	T	C	T	A	C	T	C	T	A	A	G/A	T	G/A	G	G/A	C	T	T/C	C	T	A	G	6
43	T	C	T	A/T	C	T	C	T	A	A	G	T	G	G	A	C	T	C	C	T	A	G	6
44	T	C	T	A	C	T	C	T	A	G	G	T	G	G	A	C	T	C	C	T	G	G	7
45	T	C	T	A	C	T	C	T	A	A/G	G/A	T	G/A	G	G/A	C	T	T/C	C	T	A/G	G	7



The haplotype pairs shown in Table 4 were estimated from the unphased genotypes using an extension of Clark's algorithm (Clark, A.G. (1990) *Mol Bio Evol* 7, 111-122), as described in U.S. Provisional Patent Application filed April 19, 2000 and entitled "A Method and System for Determining Haplotypes from a Collection of Polymorphisms". In this method, haplotypes are assigned  
5 directly from individuals who are homozygous at all sites or heterozygous at no more than one of the variable sites. This list of haplotypes is then used to deconvolute the unphased genotypes in the remaining (multiply heterozygous) individuals.

By following this protocol, it was determined that the Index Repository examined herein and, by extension, the general population contains the 20 human IGERA haplotypes shown in Table 5  
10 below.

Table 5. Haplotypes Observed in the IGERA Gene																							
Haplotype Number	Polymorphic Sites																						
	PS1	PS2	PS3	PS4	PS5	PS6	PS7	PS8	PS9	PS10	PS11	PS12	PS13	PS14	PS15	PS16	PS17	PS18	PS19	PS20	PS21	PS22	
1	T	C	T	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	C	T	A	G	
2	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T	C	T	A	G	
3	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	C	C	T	C	T	A	A	
4	T	C	C	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	C	T	A	G	
5	G	C	T	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	C	T	A	G	
6	T	C	T	A	C	T	C	T	A	A	G	T	G	G	A	C	T	C	C	T	A	G	
7	T	C	T	A	C	T	C	T	A	G	G	T	G	G	A	C	T	C	C	T	G	G	
8	T	C	T	A	C	T	C	T	A	A	G	T	G	G	G	C	T	T	C	T	A	G	
9	T	T	T	A	C	T	C	C	G	A	G	T	G	G	G	C	C	T	C	T	A	G	
10	T	C	T	A	C	T	C	T	A	A	A	T	A	G	G	C	T	T	C	T	A	G	
11	G	C	T	A	C	C	C	C	A	A	G	C	G	G	G	C	T	T	C	C	A	G	
12	T	C	T	A	C	T	C	C	A	A	G	T	G	G	G	C	T	T	A	T	A	G	
13	T	C	T	A	C	T	C	T	A	A	G	T	G	G	A	C	T	C	C	T	G	G	
14	T	T	T	A	C	T	C	T	A	A	G	T	G	G	G	C	C	T	C	T	A	A	
15	T	C	T	A	C	T	A	T	A	A	G	T	G	G	G	C	T	T	C	T	A	G	
16	T	C	T	A	C	T	C	C	A	A	G	T	G	A	G	C	T	T	A	T	A	G	
17	T	C	T	A	C	T	C	T	A	G	G	T	G	G	G	C	T	T	C	T	A	G	
18	T	C	T	T	C	T	C	T	A	A	G	T	G	G	A	C	T	C	C	T	A	G	
19	T	T	T	A	C	T	C	C	A	A	G	T	G	G	G	T	C	T	C	T	A	G	
20	T	T	T	A	A	T	C	T	A	A	G	T	G	G	A	C	C	C	C	T	A	G	

In view of the above, it will be seen that the several advantages of the invention are achieved and other advantageous results attained.

As various changes could be made in the above methods and compositions without departing  
5 from the scope of the invention, it is intended that all matter contained in the above description and shown in the accompanying drawings shall be interpreted as illustrative and not in a limiting sense.

All references cited in this specification, including patents and patent applications, are hereby incorporated in their entirety by reference. The discussion of references herein is intended merely to  
10 summarize the assertions made by their authors and no admission is made that any reference constitutes prior art. Applicants reserve the right to challenge the accuracy and pertinency of the cited references.

## What is Claimed is:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of:
  - (a) a first nucleotide sequence which is a polymorphic variant of a reference sequence for Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:1, and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, adenine at PS7, thymine at PS8, guanine at PS9, guanine at PS10, adenine at PS11, cytosine at PS12, adenine at PS13, adenine at PS14, adenine at PS15, thymine at PS16, thymine at PS17, cytosine at PS18, adenine at PS19, cytosine at PS20, guanine at PS21 and adenine at PS22; and
  - (b) a second nucleotide sequence which is complementary to the first nucleotide sequence.
2. The isolated polynucleotide of claim 1 which comprises a IGERA isogene.
3. The isolated polynucleotide of claim 1 which is a DNA molecule and comprises both the first and second nucleotide sequences and further comprises expression regulatory elements operably linked to the first nucleotide sequence.
4. A recombinant organism transformed or transfected with the isolated polynucleotide of claim 1, wherein the organism expresses a IGERA protein encoded by the first nucleotide sequence.
5. The recombinant organism of claim 4 which is a nonhuman transgenic animal.
6. The isolated polynucleotide of claim 1, wherein the first nucleotide sequence is a polymorphic variant of a fragment of the IGERA gene, the fragment comprising one or more polymorphisms selected from the group consisting of guanine at PS1, cytosine at PS2, cytosine at PS3, thymine at PS4, adenine at PS5, cytosine at PS6, adenine at PS7, thymine at PS8, guanine at PS9, guanine at PS10, adenine at PS11, cytosine at PS12, adenine at PS13, adenine at PS14, adenine at PS15, thymine at PS16, thymine at PS17, cytosine at PS18, adenine at PS19, cytosine at PS20, guanine at PS21 and adenine at PS22.
7. An isolated polynucleotide comprising a nucleotide sequence which is a polymorphic variant of a reference sequence for the IGERA cDNA or a fragment thereof, wherein the reference sequence comprises SEQ ID NO:2 and the polymorphic variant comprises at least one polymorphism selected from the group consisting of guanine at a position corresponding to nucleotide 251, adenine at a position corresponding to nucleotide 302, thymine at a position corresponding to nucleotide 530 and adenine at a position corresponding to nucleotide 741.
8. A recombinant organism transformed or transfected with the isolated polynucleotide of claim 7, wherein the organism expresses a Immunoglobulin E Receptor I Alpha Subunit (IGERA) protein encoded by the polymorphic variant sequence.
9. The recombinant organism of claim 8 which is a nonhuman transgenic animal.
10. An isolated polypeptide comprising an amino acid sequence which is a polymorphic variant of a

reference sequence for the IGERA protein or a fragment thereof, wherein the reference sequence comprises SEQ ID NO: 3 and the polymorphic variant comprises one or more variant amino acids selected from the group consisting of arginine at a position corresponding to amino acid position 84, asparagine at a position corresponding to amino acid position 101, methionine at a position corresponding to amino acid position 177 and lysine at a position corresponding to amino acid position 247.

11. An isolated antibody specific for and immunoreactive with the isolated polypeptide of claim 10.
12. A method for screening for drugs targeting the isolated polypeptide of claim 10 which comprises contacting the IGERA polymorphic variant with a candidate agent and assaying for binding activity.
13. A composition comprising at least one genotyping oligonucleotide for detecting a polymorphism in the Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene at a polymorphic site selected from PS1-22.
14. The composition of claim 13, wherein the genotyping oligonucleotide is an allele-specific oligonucleotide that specifically hybridizes to an allele of the IGERA gene at a region containing the polymorphic site.
15. The composition of claim 14, wherein the allele-specific oligonucleotide comprises a nucleotide sequence selected from the group consisting of of SEQ ID NOS:4-47, the complements of SEQ ID NOS: 4-47, and SEQ ID NOS:48-135.
16. The composition of claim 13, wherein the genotyping oligonucleotide is a primer-extension oligonucleotide.
17. A method for genotyping the Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene of an individual, comprising determining for the two copies of the IGERA gene present in the individual the identity of the nucleotide pair at one or more polymorphic sites (PS) selected from PS1-22.
18. The method of claim 17, wherein the determining step comprises:
  - (a) isolating from the individual a nucleic acid mixture comprising both copies of the IGERA gene, or a fragment thereof, that are present in the individual;
  - (b) amplifying from the nucleic acid mixture a target region containing at least one of the polymorphic sites;
  - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
  - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and
  - (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
19. A method for haplotyping the Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene of an

individual which comprises determining, for one copy of the IGERA gene present in the individual, the identity of the nucleotide at one or more polymorphic sites (PS) selected from PS1-22.

20. The method of claim 19, wherein the determining step comprises
- (a) isolating from the individual a nucleic acid molecule containing only one of the two copies of the IGERA gene, or a fragment thereof, that is present in the individual;
  - (b) amplifying from the nucleic acid molecule a target region containing at least one of the polymorphic sites;
  - (c) hybridizing a primer extension oligonucleotide to one allele of the amplified target region;
  - (d) performing a nucleic acid template-dependent, primer extension reaction on the hybridized genotyping oligonucleotide in the presence of at least two different terminators of the reaction, wherein said terminators are complementary to the alternative nucleotides present at the polymorphic site; and
  - (e) detecting the presence and identity of the terminator in the extended genotyping oligonucleotide.
21. A method for predicting a haplotype pair for the Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene of an individual comprising:
- (a) identifying an IGERA genotype for the individual at two or more of polymorphic sites selected from PS1-22;
  - (b) enumerating all possible haplotype pairs which are consistent with the genotype;
  - (c) accessing data containing the IGERA haplotype pairs determined in a reference population; and
  - (d) assigning a haplotype pair to the individual that is consistent with the data.
22. A method for identifying an association between a trait and at least one genotype or haplotype of the Immunoglobulin E Receptor I Alpha Subunit gene which comprises comparing the frequency of the genotype or haplotype in a population exhibiting the trait with the frequency of the genotype or haplotype in a reference population, wherein the genotype or haplotype comprises a nucleotide pair or nucleotide located at one or more polymorphic sites selected from PS1-22, wherein a higher frequency of the genotype or haplotype in the trait population than in the reference population indicates the trait is associated with the genotype or haplotype.
23. The method of claim 22, wherein the haplotype is selected from haplotype numbers 1-20 shown in Table 5.
24. The method of claim 23, wherein the trait is a clinical response to a drug targeting IGERA .
25. A computer system for storing and analyzing polymorphism data for the Immunoglobulin E Receptor I Alpha Subunit gene, comprising:
- (a) a central processing unit (CPU);
  - (b) a communication interface;

- (c) a display device;
  - (d) an input device; and
  - (e) a database containing the polymorphism data;
- wherein the polymorphism data comprises the genotypes and haplotype pairs shown in Table 4 and the haplotypes shown in Table 5.
- 10
26. A genome anthology for the Immunoglobulin E Receptor I Alpha Subunit (IGERA) gene which comprises IGERA isogenes defined by haplotypes 1-20 shown in Table 5.

1/6

## POLYMORPHISMS IN THE IGERA GENE (Accession No. L14075)

GATCTTCATG	TGGAATGACT	GGTTTCATTC	AATAGACTTA	ATTCAGCAGT	
CTGTGGGGAA	GAGCAAGGTA	TGATAGAATG	GTTCCCTCAAG	TGCTTCAGAT	100
GTGAAGTGGG	TTTAAATATA	CTGTCCCTGT	CTTCTTCAGA	GTTTTGGTAA	
AGATAAAATA	GGACACTCAT	TTAAAAGCAA	TCTTTGCAAA	TGACAAGCCA	200
CTATAGACAT	TAATAGAGTT	TTCATTTCCA	GTATTATCAT	TAATATCAGA	
TCCTGGAAGA	AGGTTGAGCC	TTGACCTAGA	GCAAAAAAAC	AGAAGAATTA	300
GTAAAGGAAT	CCTGGAGAAA	GCCCCTGCTG	TGTATTTAAA	GGAGAAAGGG	
AGATCATGTT	GGGAAATTAT	AATATTAAAA	GTAAACAAAA	GCTAGGAAGT	400
AAAATTAAT	AAATTATATG	GCCTAGATCC	CCATAAGTAA	TGGTTTAACT	
TCTGCCTTCC	TGTGTTCTGA	GCCAGATTAG	GGCACAGTAG	AGAAAGAGGA	500
GTCTCTGAAA	ATGTTTCCAA	TTTCGCTGGT	CAGACAGCGG	ATCATCAGTG	
AATCAGATGA	AAATTTGTGG	ATTTATGCAC	TAAGTGATCA	GCAGGAAATT	600
AAACAAGAAA	AGCGTTGGTA	GCTCTGGTGA	ATCCCAAAAG	AATTTGGCAG	
TTGCTAGCCA	TGCTCCTGAA	TATGTATAAA	CAGTACATCA	TATGACTAAG	700
AGTTTGACTT	AGGGGTTAGA	TTTTATGTGT	TTGAACCCCA	AATTAGTTAT	
TTAATAGTTG	GCACCCCAAA	ACAAGTTACT	TAACCTCACT	AAGATTCAGT	800
TTTCCTGTTT	ATAAAATGTA	GATAGTGATA	GTATGTACTT	TATAGGATTA	
TTGTGAAAAA	TAAATGAAAT	ATCAGATTTA	TTTAGGATAA	CACCTGGCAT	900
	G				
ATGTTTGGTA	TTCAGTAATT	AGTTGCTGCT	GTTTTATTCT	GCTCTCCCTT	
			C		
GCATCCCACT	TTTCTAAGTT	GTAAACTAAA	TAGTTGTACA	CAGATTGACA	1000
GATTAAAGAAA	GGCTTGTGAT	TGTGCTAGAC	CTATGCCTCT	CTCTCACCAG	
ATTCCAGGTG	TATATGTGGA	GGTGGGATAG	GGAGTGGAGT	AAGTGGGTAA	1100
ATATTAAATT	GCCCAGTTGG	GCACCATCCT	GAATATTATC	TCTAAAGAAA	
GAAGCAAAAC	CAGGCACAGC	TGATGGGTTA	ACCAGATATG	ATACAGAAAA	1200
			C	T	
CATTTCCCTC	TGCTTTTTTG	TTTTAAGCCT	ATATTTGAAG	CCTTAGATCT	
CTCCAGCACA	GTAAGCACCA	GGAGTCCATG	AAGAAGATGG	CTCCTGCCAT	1300
	[exon 1: 1287..				
GGAATCCCCT	ACTCTACTGT	GTGTAGCCTT	ACTGTTCTTC	GGTAAGTAGA	
	..1341]				
GATTCAATTA	CCCCTCCCAG	GGAGGCCCAA	ATGAATTTGG	GGAGCAGCTG	1400
	A				
GGGTAGGAAC	CTTTACTGTG	GGTGGTGACT	TTTTCTAGGA	CATGTGCAAA	
CTATTGGGCA	TTTCCCAGGG	ACTCTGTAGT	GGAGCCAAGC	TAGAAAGCAG	1500
AGGCAAGTGG	GCTGAGCAAC	ACCTAAGGAG	GAAGCCAGAC	TGAAAGCTTG	
GTTCCCTTGA	TTTGCTCTGG	CATCTTCCAG	AGTGCAAATT	TCCTACCAAG	1600
GTAATGAGGG	TAGAGGAGAG	AAAGAAGCTC	TTTCTTCCCC	TGATTCTCAT	
TCCTGAAAAG	ACGTTTGGTC	CTTAAAATTC	CATGGATGTA	GATCTTATCC	1700
CCACACCCAG	ATTCTAGTCC	TCTGGAGATA	AAGAAGACTG	CTGGACACTA	
ATGTATCCTC	TCTGGACTTT	TGCAGCTCCA	GATGGCGTGT	TAGCAGGTGA	1800
	C	A			
	[exon 2: 1776..	..1796]			
GTCCTCTGTT	CTTGTTCCCT	TGGTGTATCA	ACATGTCTGG	GCATTGCTTT	
CCTCTCACTA	TTTTCTTCGT	CCCATCACTT	CTGCTTTCTA	ATGAGCATGA	1900
			T		
ATCTGTTTCT	TGGCCAGACT	ACTTTCCCTC	TCCACCTTGC	CTTGTCTTTC	
TTTTTTTCCC	TGATTCATTG	CATTCTCTCA	AGTCATTCTC	TCCTCTGTTT	2000
TAGTCAATAA	CCATGTCTGT	TGCACATATA	CATGTCTCAT	TCTCTCTCCT	
AGACACTTTG	GCATGATCTC	GCTCAATAAT	TACATTATTA	TTATTATTGC	2100

FIGURE 1A



2/6

CATTTTATAA	TTGAGGATGC	TGAAACTCAG	TGATTTTCTG	GTGGTTACAT	
GGCTAAGGAA	CTGGATTTCA	ACGTAAGTTC	CTTGGATCTA	AGTCCAGTTC	2200
TCTTCTGACT	ATATCACCCT	TTTGTATCA	CCATGTATCT	ACTTCTTTGG	
TCTCTGTTCA	AATTTGCACT	ACATCCCCTT	GTTCCAGGAA	GCCATTCAAG	2300
ACTGACTTTC	TTAGTGCCTC	TCACTACTTT	CTGGAAGTGA	CATATGTTTT	
TCACTCTGTA	TATACTTACA	ATTAAATAGT	CATAAATATT	CAGAGCTTGG	2400
AGAAACCTTA	TATTTTCATCC	AGTCCAGTAA	ATTTATCCAT	CCATAATTCA	
CTCATTTCATT	CACATAATAA	ATATTTAATG	TAACAATGGT	TGAACATGGC	2500
AGACAGTGTT	TCTACCTCAA	AAGAGATTGC	AGTCCTCATT	TACAGATACT	
GAATTGAAAT	TAACAGAAGT	AGAGTGAGTC	AGCTCAAATC	ACATAGTGAA	2600
TTGGTTTCTT	TGTTTTTAAA	TCTCCTGCAT	ATGTGTCCTG	TCTTTCTCCC	
TGTGTTGGGC	GTTCCCTGGG	GCACCAATAC	TAATTTCTCC	TTCCCCTAGA	2700
AATCAAAACA	GGGTCTTATC	ACCAACAGAA	TAAGGACAGG	TTGACCACTG	
G					
ATTGTCAGAA	TATTGCTTCG	TTTGTACTTT	TAAGCCTAGA	CAGTTTTTCAA	2800
TGACTTTTTT	TCTCTCTACA	TGTCTTTTCA	TATTTTTTATC	TTCTTGAAGT	
[exon 3: 2850..					
CCCTCAGAAA	CCTAAGGTCT	CCTTGAACCC	TCCATGGAAT	AGAATATTTA	2900
AAGGAGAGAA	TGTGACTCTT	ACATGTAATG	GGAACAATTT	CTTTGAAGTC	
AGTTCCACCA	AATGGTTCCA	CAATGGCAGC	CTTTCAGAAG	AGACAAATTC	3000
AAGTTTGAAT	ATTGTGAATG	CCAAATTTGA	AGACAGTGGA	GAATACAAAT	
G					
GTCAGCACCA	ACAAGTTAAT	GAGAGTGAAC	CTGTGTACCT	GGAAGTCTTC	3100
A					
AGTGGAAGT	TCCAGGGATA	TGGAATATACA	GATCTCTCAT	GTGAGGGATG	
..3104]					
GCTCATCTGA	AGATGGGAAA	AAACAGGTTA	TTCCAAGGGT	TAGGACACCA	3200
GAGTGGGATT	CAAGGCCTCT	CATTTTAAAG	ACCCCTGCAT	TGGCTGGGCA	
C					
CAGTGGCTCA	CGCCTGTAAT	CCCAGCACTT	TGGGAGGCTG	AGGCAGGTGG	3300
A					
ATCACGAGGT	CAGGAGATCG	AGACCATCCG	GCTAACATGG	TGAAACCCCA	
A					
TCTCTGCTAA	AAAATATATA	TATATAAAAT	TAGCCGGGCG	TAGTGGTGGG	3400
CACCTGTAGT	CCCAGGTACT	CGGGAGGCTG	AGGCAGGAGA	ATGGTGTGAA	
CCCAGGAGGT	GGAGGTGCA	GTGAGCTGAG	ATCACGCCAC	TGCCCTCCAG	3500
CCTGGGCTAC	AGAGCAAGAC	TCCGTCTCAA	AAAATAAATA	AATAAATAAA	
AAAGACCCCT	GCATCTCTTT	TCTTCTACCC	CCTTCCCTTT	TGATTACTTG	3600
TATGCCTTCT	TTCAATATTC	TAGTCATCTC	TCAATATTAT	TCCTCCACCC	
TATTTTCTCT	TATCTTTTCT	GCCTAGATTG	AGGTATATAT	TATGTGGTCA	3700
AACAGCATGA	CATATATGTG	AACATTTCAA	AGAGCTGTGT	ATCTGGAATA	
GGATCAAAAG	GTTTGACTTA	AAGTTTTGCT	CTGCATAATC	CATATGGCAG	3800
GACCTGAATA	TTAGGTTGTA	CTCTTCGTTA	TGAAACATAT	CTGGGTACAT	
TTCCTTATGT	CCTCTGTTGT	TACTTAAGAA	CACATATTTT	ATGCTTGTTT	3900
CATTTTTATC	ACTCCTACTG	CCAACAAATA	GCATAGCATG	CTTAGGCACA	
TGTGGCTTAA	TTAGCAAATG	TTGAATAAAC	AAATTAATGA	TTTTGAATAG	4000
TGACCAATAG	GTCTCTTTTA	TACTCTATAT	TTTTCTCTTG	AGTGAAAAAA	
AATGTTTCAA	CCTCCATATG	TAAATTCCAA	ACACAAACTA	AAGCAATGTA	4100
GAATAGCTTC	TTTATTCCCT	GGAGTAGGTT	CTAGAGAAGT	CCTAAAGGAT	
TGGTCCTAAA	TTAATTATGC	TTATTATGCT	AGCGATATTT	CCTTTCAAAA	4200
TTCTCCTTTA	ATGAATGCTT	TTTAATTTTT	ACAAAAGCAT	TAACCATAGA	
ATGTGATTCT	TGTCTTTCAC	TGACTCATTA	GTGACAAATA	TTTGTTGAGT	4300
ACCTACCAAC	TCCTAAGTAT	TGCTACCAAC	TCCTAAATAC	TGTGTTGGGC	
ATTCAGAATA	GAATGTAGAA	CTAGACAGGG	TCCCTGACTT	CTTGAGGCAC	4400

FIGURE 1B

3/6

AGAGCAGTAT	GGGAAGAGGA	CATTAAATAA	AGAATTACAT	AAGTAATTAA	
TTTAAATTAT	ACATGTTTTG	AAGAAGTTTT	TTTTTGACAA	CTATAATTAA	4500
CACTAGAACT	GGGAAGTTTC	TATAAGGTAA	GAGAGGACAA	AATAGACACT	
CTCCTAAGCT	AAAATTCCCA	AGAAAGACTG	TTTATTTTCC	CCTAACTAAC	4600
TAGAACTAGC	AACAGAAGAT	CTGAAAGGAA	TTCTGGCTTT	CAAGTGTTC	
ATGTATGGAC	TCATCAGGGA	GGTCCGAGAG	GCTTTGTGGC	CCCAGACTGA	4700
CTTTTCAGGA	GGGGAAAGGA	TTTATCAATA	CACAAGACAG	GCTCTAAGCA	
TTATTTTGTG	CCCTTTAAAA	ATCCACTTTA	TGAGCCAAAA	AGTGAGTTAA	4800
TGATAATTCA	TAGTTTCTGA	CACATGCTCT	ATGCGTGCCT	CTCTTTTCTC	
A					
TATTCATTCT	CTCTCTCTTC	ATTTATTGTT	AAATAAATAA	TGTAATGAAT	4900
GTTCTTCAGA	CTGGCTGCTC	CTTCAGGCCT	CTGCTGAGGT	GGTGATGGAG	
[exon 4: 4910..					
GGCCAGCCCC	TCTTCCTCAG	GTGCCATCCT	TGGAGGAACT	GGGATGTGTA	5000
CAAGGTGATC	TATTATAAGG	ATGGTGAAGC	TCTCAAGTAC	TGGTATGAGA	
ACCACAACAT	CTCCATTACA	AATGCCACAG	TTGAAGACAG	TGGAACCTAC	5100
TACTGTACGG	GCAAAGTGTG	GCAGCTGGAC	TATGAGTCTG	AGCCCCCTCAA	
T					
CATTACTGTA	ATAAAAGGTG	AGTTGGTAAA	GGAAAGGAAA	AGCATCCATA	5200
..5167]					
GCAGGGGAAG	GAAGAGAGAA	CTTCTGAGCC	TGAGCAGTTG	CAGCTTGTAAG	
AAGGGGGGCA	CCTGTGATAC	ACTGGAAAGC	CTACCAGACT	TGCAATGAGG	5300
T					
AGACCTGGGT	GATAGTATAT	ATCTCAATCT	CTGTTTCAAA	GCCTTGACTT	
GTTAAATGGT	GATAGTAATA	CCTGCTTGCA	CTATGAAATT	TTTATGAAGA	5400
C					
TTAATGTGGT	AATATTTGTG	AAATGACTTT	GTAAACTGTT	AAGCACTACC	
CAAGCATAAC	AGATTGTGAT	TACTATTTTG	ATCTCAAAGT	CATCTGTTGC	5500
TCCTGGGGGA	ACACTTATAT	TTATCAAATT	GAAAAAAAGT	TTCAAAGTTG	
AATGAAGAAA	GGATATAAAG	AGCTTGAGGA	GCCCATTTCCA	GCTTAGGAGG	5600
GCTGGGAAAG	GAAACCAGCA	AGTCAGTAAG	CTGTGTGCCT	GTGTATTGAG	
GGAGGAGGGA	ATGGACTTGA	TATGGAGAGG	GTAGGGAGGT	GGACTGCCTC	5700
TATGGCCTGT	AAGAAAAACT	GCTCTCTCCA	AACTCTTTAT	AAGAGAGGGA	
GCCTGTGAAG	TATTCACTTT	TGAAGGAGAA	AGTTAGACTT	TTCTTTCACA	5800
CACTTTGTAC	ATAATAATGT	TTAAAAAAGC	ATGAGGTCAA	AATACATAAT	
TAAGTCCTAG	CAGTTCTCTG	TTAACTAATT	TGAGACTGAA	GTGCTATGTA	5900
CTTGTCTCTA	GGCTTCCAGT	ATCTTCATCT	GTAAAACAGA	ATATTTGGTC	
TAGATTCCAT	TAGAATCATT	TGATAACTTA	AAAAATATAT	TGATGCTCAT	6000
GTCTCATTTC	TTGAGATTCT	GATTTAATTG	GTTTGGGGTG	CAGCCTGGGT	
ATACGTATTT	TTCATAGGTC	TTTCACATAA	TGGTAATGGG	TAGCCAATAT	6100
TGAGAATCAC	TTGTCTAGGT	GATCTTTAAA	TGATTTCTGG	ATGTAATATT	
CTGAGGCTCT	ATAATTTGAG	ACTAATCACA	AAAATCGGTA	CAGTTTATAA	6200
ACAGACTAAC	AGAACCACAA	AATAATAGAA	TTGGAAGGCA	ATTTAACTAG	
TGCAATTTCT	TCATTTTGCC	TAACAGGCAT	GTAAGAAATG	ATGATTGATT	6300
GAGTAATAGG	CATTGATGAC	CCCTGTCCTC	ACTTTGTCCC	CTTTCCACCC	
CTTAATTATA	TGTGAATTCT	GGTCTTGTC	TTTCGAATAA	GGGGTTTATC	6400
TTTCCTATTG	TCTTCCCCTC	TGGGCACGGC	ACACTGGCTA	CTGGAGTTAA	
GAGGAAATGC	TTAGGACTCC	CTGTGGCTCC	AGGGAGCACC	AACAGAGCAA	6500
CTCAACCTAG	TGTTAATCTG	AGTGTTTTCT	CTGTGCTTCT	GGATGCCACA	
TCACGCTAAA	AATGAAGGAC	AAAGCTTGGT	CTTCTCTTA	GGGAGGATGA	6600
AACTCTGAAC	CTCATTTTTT	AGTTCCCAAG	ATGAATTATG	TTTCTCATTG	
CATCTGTGTT	CCACTACAGC	TCCGCGTGAG	AAGTACTGGC	TACAATTTTT	6700
[exon 5: 6670..					
TATCCCATTTG	TTGGTGGTGA	TTCTGTTTGC	TGTGGACACA	GGATTATTTA	

FIGURE 1C

4/6

TCTCAACTCA	GCAGCAGGTC	ACATTTCTCT	TGAAGATTAA	GAGAACCAGG	6800
AAAGGCTTCA	GACTTCTGAA	CCCACATCCT	AAGCCAAACC	CCAAAAACAA	
		A			
CTGATATAAT	TACTCAAGAA	ATATTTGCAA	CATTAGTTTT	TTTCCAGCAT	6900
	..6854]				
CAGCAATTGC	TACTCAATTG	TCAAACACAG	CTTGCAATAT	ACATAGAAAC	
	C		G		
GTCTGTGCTC	AAGGATTTAT	AGAAATGCTT	CATTAAACTG	AGTGAAACTG	7000
				A	
GTTAAGTGGC	ATGTAATAGT	AAGTGCTCAA	TTAACATTGG	TTGAATAAAT	
GAGAGAATGA	ATAGATTCAT	TTATTAGCAT	TTGTAAAAGA	GATGTTCAAT	7100
TTCAATAAAA	TAAATATAAA	ACCATGTAAC	AGAATGCTTC	TGAGTA	7146

FIGURE 1D

5/6

## POLYMORPHISMS IN THE CODING SEQUENCE OF IGERA

ATGGCTCCTG	CCATGGAATC	CCCTACTCTA	CTGTGTGTAG	CCTTACTGTT	
CTTCGCTCCA	GATGGCGTGT	TAGCAGTCCC	TCAGAAACCT	AAGGTCTCCT	100
TGAACCCTCC	ATGGAATAGA	ATATTTAAAG	GAGAGAATGT	GACTCTTACA	
TGTAATGGGA	ACAATTTCTT	TGAAGTCAGT	TCCACCAAAT	GGTTCACAA	200
TGGCAGCCTT	TCAGAAGAGA	CAAATTCAAG	TTTGAATATT	GTGAATGCCA	
AATTGAAGA	CAGTGGAGAA	TACAAATGTC	AGCACCAACA	AGTTAATGAG	300
G					
AGTGAACCTG	TGTACCTGGA	AGTCTTCAGT	GACTGGCTGC	TCCTTCAGGC	
A					
CTCTGCTGAG	GTGGTGATGG	AGGGCCAGCC	CCTCTTCCTC	AGGTGCCATG	400
GTTGGAGGAA	CTGGGATGTG	TACAAGGTGA	TCTATTATAA	GGATGGTGAA	
GCTCTCAAGT	ACTGGTATGA	GAACCACAAC	ATCTCCATTA	CAAATGCCAC	500
AGTTGAAGAC	AGTGGAACCT	ACTACTGTAC	GGGCAAAGTG	TGGCAGCTGG	
		T			
ACTATGAGTC	TGAGCCCCTC	AACATTACTG	TAATAAAAGC	TCCGCGTGAG	600
AAGTACTGGC	TACAATTTTT	TATCCCATTG	TTGGTGGTGA	TTCTGTTTGC	
TGTGGACACA	GGATTATTTA	TCTCAACTCA	GCAGCAGGTC	ACATTTCTCT	700
TGAAGATTAA	GAGAACCAGG	AAAGGCTTCA	GAATTCTGAA	CCCACATCCT	
			A		
AAGCCAAACC	CCAAAAACAA	CTGA			774

FIGURE 2

6/6

## ISOFORMS OF THE IGERA PROTEIN

MAPAMESPTL	LCVALLFFAP	DGVLAVPQKP	KVSLNPPWNR	IFKGENVTLT	
CNGNNFFEVS	STKWFFHNGSL	SEETNSSLNI	VNAKFEDSGE	YKCQHQQVNE	100
			R		
SEPVYLEVFS	DWLLLQASAE	VVMEGQPLFL	RCHGWRNWDV	YKVIYYKDGE	
N					
ALKYWYENHN	ISITNATVED	SGTYICTGKV	WQLDYESEPL	NITVIKAPRE	200
		M			
KYWLQFFIPL	LVVILFAVDT	GLFISTQQQV	TELLKIKRTR	KGFRLLNPHP	
				K	
KPNPKNN					257

FIGURE 3

## SEQUENCE LISTING

<110> Genaissance Pharmaceuticals

Denton, R. Rex  
Nandabalan, Krishnan  
Kliem, Stephanie  
Chew, Anne  
Duda, Amy  
Lanz, Elizabeth

<120> Drug Target Isogenes: Polymorphisms in the  
Immunoglobulin E Receptor I Alpha Subunit Gene

<130> MWH-0007PCT IGERA/FCER1A

<140> TBA

<141> 2000-08-02

<150> 60/147,860

<151> 1999-08-09

<160> 219

<170> PatentIn Ver. 2.1

<210> 1

<211> 7146

<212> DNA

<213> Homo sapiens

<400> 1

gatcttcatg tggaatgact ggtttcattc aatagactta attcagcagt ctgtggggaa 60  
gagcaaggta tgatagaatg gttcctcaag tgcttcagat gtgaagtggg tttaaatata 120  
ctgtccctgt cttcttcaga gttttggtta agataaaaata ggacactcat ttaaaagcaa 180  
tctttgcaaa tgacaagcca ctatagacat taatagagtt ttcattttcca gtattatcat 240  
taatatcaga tcctggaaga aggttgagcc ttgacctaga gcaaaaaaac agaagaatta 300  
gtaaaaggaat cctggagaaa gccctgctg tgtattttaa ggagaaaggg agatcatggt 360  
gggaaattat aatattaaaa gtaaacaaaa gctaggaagt aaaataaaat aaatttatatg 420  
gcctagatcc ccataagtaa tggtttaact tctgccttcc tgtgttctga gccagattag 480  
ggcacagtag agaaagagga gtctctgaaa atgtttccaa tttcgctggt cagacagcgg 540  
atcatcagtg aatcagatga aaatttggtg atttatgcac taactgatca gcaggaaatt 600  
aaacaagaaa agcgttggtg gctctggtga atcccaaaag aatttggcag ttgctagcca 660  
tgctcctgaa tatgtataaa cagtacatca tatgactaag agtttgactt aggggttaga 720  
ttttatgtgt ttgaacccca aattagttat ttaatagttg gcaccccaaa acaagttact 780  
taacctcact aagattcagt tttcctgttt ataaaatgta gatagtata gtatgtactt 840  
tataggatta ttgtgaaaaa taaatgaaat atcagattta tttaggataa cacctggcat 900  
atgtttggta ttcagtaatt agttgctgct gttttattct gctctccctt gcacccact 960  
tttctaagtt gtaactaaa tagttgtaca cagattgaca gattaagaaa ggcttgatgat 1020

tgtgctagac ctatgcctct ctctcaccag attccagggtg tatatgtgga ggtgggatatg 1080  
 ggagtggagt aagtgggtaa atattaaatt gccagttgg gcaccatcct gaattattatc 1140  
 tctaaagaaa gaagcaaaac caggcacagc tgatgggtta accagatatg atacagaaaa 1200  
 catttccttc tgcttttttg ttttaagcct atatttgaag ccttagatct ctccagcaca 1260  
 gtaagcacca ggagtccatg aagaagatgg ctccctgccat ggaatcccct actctactgt 1320  
 gtgtagcctt actgttcttc ggtaagtaga gattcaatta cccctcccag ggaggcccaa 1380  
 atgaatttgg ggagcagctg gggtaggaac ctttactgtg ggtgggtgact ttttctagga 1440  
 catgtgcaaa ctattgggca tttcccaggg actctgtagt ggagccaagc tagaaagcag 1500  
 aggcaagtgg gctgagcaac acctaaggag gaagccagac tgaaagcttg gttccttgca 1560  
 tttgctctgg catcttccag agtgcaaatt tcttaccag gtaatgaggg tagaggagag 1620  
 aaagaagctc tttcttcccc tgattctcat tctgaaaag acggttggtc cttaaaattc 1680  
 catggatgta gatcttatcc ccacaccag attctagtcc tctggagata aagaagactg 1740  
 ctggacacta atgtatcctc tctggacttt tgcagctcca gatggcgtgt tagcagggtga 1800  
 gtccctctgtt cttgttccct tgggttatca acatgtctgg gcattgcttt cctctcacta 1860  
 ttttcttctg cccatcactt ctgctttcta atgagcatga atctgttctt tggccagact 1920  
 actttccctc tccaccttgc cttgtctttc ttttttccc tgattcattg cattctctca 1980  
 agtcattctc tctctgttt tagtcaataa ccatgtctgt tgcacatata catgtctcat 2040  
 tctctctcct agacactttg gcatgatctc gctcaataat tacattatta ttattattgc 2100  
 cattttataa ttgaggatgc tgaaactcag tgattttctg gtggttacat ggctaaggaa 2160  
 ctggatttca acgtaagtcc cttggatcta agtccagttc tcttctgact atatcacctt 2220  
 tttgttatca ccatgtatct acttcttttg tctctgttca aatttgcact acatcccctt 2280  
 gttccaggaa gccattcaag actgactttc ttagtgcctc tcaactactt ctggaactga 2340  
 catatgtttt tcaactctgta tatacttaca attaaatagt cataaatatt cagagcttgg 2400  
 agaaacctta tatttcatcc agtccagtaa atttatccat ccataattca ctcatcatt 2460  
 cacataataa atatttaatg taacaatggt tgaacatggc agacagtgtt tctacctcaa 2520  
 aagagattgc agtccctatt tacagatact gaattgaaat taacagaagt agagtgaagtc 2580  
 agctcaaadc acatagttaa ttggtttctt tgttttttaa tctcctgcat atgtgtcctg 2640  
 tctttctccc tgtgttgggc gttccctggg gcaccaatac taatttctcc tccccctaga 2700  
 aatcaaaaaca gggctttatc accaacagaa taaggacagg ttgaccactg attgtcagaa 2760  
 tattgcttctg tttgtacttt taagcctaga cagttttcaa tgactttttt tctctctaca 2820  
 tgtcttttca tatttttatc ttcttgaagt cctcagaaa cctaaggctc ccttgaaccc 2880  
 tccatggaat agaattttta aaggagagaa tgtgactctt acatgtaatg ggaacaattt 2940  
 ctttgaagtc agttccacca aatggttcca caatggcagc ctttcagaag agacaaattc 3000  
 aagtttgaat attgtgaatg ccaaatttga agacagtgga gaatacaaat gtcagcacca 3060  
 acaagttaat gagagtgaac ctgtgtacct ggaagtcttc agtggtaagt tccagggata 3120  
 tggaaataca gatctctcat gtgagggtat gctcatctga agatgggaaa aaacaggtta 3180  
 ttccaagggt taggacacca gagtgggatt caaggcctct cattttttaag acccctgcat 3240  
 tggctgggca cagtggctca cgctgtaat cccagcactt tgggaggctg aggcagggtg 3300  
 atcacgaggt caggagatcg agaccatccg gctaacatgg tgaaaccca tctctgctaa 3360  
 aaaatatata tatataaaat tagccgggag tagtgggtggg cacctgtagt cccagggtact 3420  
 cgggaggctg aggcaggaga atggtgtgaa cccaggaggt ggagggttga gtgagctgag 3480  
 atcacgccac tgccctccag cctgggctac agagcaagac tccgtctcaa aaaataaata 3540  
 aataaataaa aaagaccctt gcatctcttt tcttctaccc ccttcccttt tgattacttg 3600  
 tatgccttct ttcaatatc tagtcatctc tcaatattat tcttccacc tatttctctc 3660  
 tatcttttct gcctagattc aggtatatat tatgtgttca aacagcatga catatatgtg 3720  
 aacatttcaa agagtgtgt atctggaata ggatcaaaag gtttgactta aagttttgct 3780  
 ctgcataatc catatggcag gacctgaata ttaggttgta ctcttcgtta tgaaacatat 3840  
 ctgggtacat ttcttatgt cctctgttgt tacttaagaa cacatatttc atgcttgttt 3900

cttttttatc actcctactg ccaacaaata gcatagcatg cttaggcaca tgtggcttaa 3960  
 ttagcaaatg ttgaataaac aaattaatga ttttgaatag tgaccaatag gtctctttta 4020  
 tactctatat ttttctcttg agtgaaaaaa aatgtttcaa cctccatag taaattccaa 4080  
 acacaaacta aagcaatgta gaatagcttc tttattccct ggagtaggtt ctagagaagt 4140  
 cctaaaggat tggctcctaaa ttaattatgc ttattatgct agcgatattt cctttcaaaa 4200  
 ttctccttta atgaatgctt ttttaattttt acaaaagcat taaccataga atgtgattct 4260  
 tgtctttcac tgactcatta gtgacaaata tttgttgagt acctaccaac tcctaagtat 4320  
 tgctaccaac tcctaaatac tgtgttgagg attcagaata gaatgtagaa ctgacaggg 4380  
 tccctgactt cttggagcac agagcagtat gggagagga cattaaataa agaattacat 4440  
 aagtaattaa tttaaattat acatgttttg aagaagttt tttttgacaa ctataattaa 4500  
 cactagaact gggaggttct tataaggtaa gagaggacaa aatagacact ctcttaagct 4560  
 aaaattccca agaaagactg tttattttcc cctaaactaac tagaactagc aacagaagat 4620  
 ctgcnaggaa ttctggcttt caagtgttc atgtatggac tcatcagggg ggtccgagag 4680  
 gctttgtggc ccagactga cttttcagga ggggaaagga tttatcaata cacaagacag 4740  
 gctctaagca ttattttgtg ccctttaaaa atccacttta tgagccaaaa agtgagttaa 4800  
 tgataattca tagtttctga cacatgctct atgcgtggct ctcttttctc tattcattct 4860  
 ctctctcttc atttattggt aaataaataa tgtaataaat gttcttcaga ctggctgctc 4920  
 cttcaggcct ctgctgaggt ggtgatggag ggcagcccc tcttctcag gtgccatggt 4980  
 tggaggaact gggatgtgta caaggtgatc tattataagg atggtgaagc tctcaagtac 5040  
 tggatgaga accacaacat ctccattaca aatgccacag ttgaagacag tggaacctac 5100  
 tactgtacgg gcaaagtgtg gcagctggac tatgagtctg agcccctcaa cattactgta 5160  
 ataaaagggt agttggtaaa ggaaaggaaa agcatccata gcaggggaag gaagagagaa 5220  
 cttctgagcc tgagcagttg cagctttagt aaggggggca cctgtgatac actggaaagc 5280  
 ctaccagact tgcaatgagg agacctgggt gatagtatat atctcaatct ctgtttcaaa 5340  
 gccttgactt gttaaagggt gatagtaata cctgcttgca ctatgaaatt tttatgaaga 5400  
 ttaatgtggt aatatttgtg aaatgacttt gtaaaactgt aagcactacc caagcataac 5460  
 agattgtgat tactattttg atctcaaagt catctgttgc tctggggga acacttatat 5520  
 ttatcaaatt gaaaaaaagt ttcaaagttg aatgaagaaa ggatataaag agcttgagga 5580  
 gccattcca gcttaggagg gctgggaaag gaaaccagca agtcagtaag ctgtgtgcct 5640  
 gtgtattgag ggaggaggga atggacttga tatggagagg gtagggaggt ggactgcctc 5700  
 tatggcctgt aagaaaaact gctctctcca aactctttat aagagaggga gcctgtgaag 5760  
 tattcacttt tgaaggagaa agttagactt ttccttcaca cactttgtac ataataatgt 5820  
 ttaaaaaagc atgagggtcaa aatacataat taagtcctag cagttctctg ttaactaatt 5880  
 tgagactgaa gtgctatgta cttgtctcta ggcttcagat atcttcatct gtaaaacaga 5940  
 atatttggtc tagattccat tagaatcatt tgataactta aanaatatat tgatgctcat 6000  
 gtctcatttc ttgagattct gatttaattg gtttggggtg cagcctgggt atacgtattt 6060  
 ttcataggtc tttcacataa tggtaatggg tagccaatat tgagaatcac ttgtctaggt 6120  
 gatctttaaa tgatttctgg atgtaattt ctgaggctct ataatttgag actaatcaca 6180  
 aaaatcggtg cagtttataa acagactaac agaaccacaa aataatagaa ttggaaggca 6240  
 atttaactag tgcaatttct tcattttgcc taacaggcat gtaagaaatg atgattgatt 6300  
 gagtaatagg cattgatgac ccctgtcctc actttgtccc ctttccaccc cttaattata 6360  
 tgtgaattct ggtcttgctc tttcgaataa ggggtttatc tttcctattg tcttcccctc 6420  
 tgggcacggc aactggcta ctggagttaa gaggaatgc ttaggactcc ctgtggctcc 6480  
 agggagcacc aacagagcaa ctcaacctag tgtaaatctg agtgttttct ctgtgcttct 6540  
 ggatgccaca tcacgctaaa aatgaaggac aaagcttggt ctttctctta yggaggatga 6600  
 aactctgaac ctcattttct agttcccaag atgaattatg tttctcattg catctgtgtt 6660  
 ccactacagc tccgctgag aagtactggc tacaattttt tatcccattg ttggtggtga 6720  
 ttctgtttgc tgtggacaca ggattattta tctcaactca gcagcaggtc acatttctct 6780



tgaagattaa gagaaccagg aaaggcttca gacttctgaa cccacatcct aagccaaacc 6840  
 ccaaaaaaaa ctgatataat tactcaagaa atatttgcaa cattagtttt tttccagcat 6900  
 cagcaattgc tactcaattg tcaaacacag cttgcaatat acatagaaac gtctgtgctc 6960  
 aaggatttat agaaatgctt cattaaactg agtgaaactg gttaagtggc atgtaatagt 7020  
 aagtgtctcaa ttaacattgg ttgaataaat gagagaatga atagattcat ttattagcat 7080  
 ttgtaaaaga gatgttcaat ttcaataaaa taaatataaa accatgtaac agaattgcttc 7140  
 tgagta 7146

<210> 2

<211> 774

<212> DNA

<213> Homo sapiens

<400> 2

atggctcctg ccatggaatc ccctactcta ctgtgtgtag ccttactgtt cttegtctcca 60  
 gatggcgtgt tagcagtcctc tcagaaacct aagggtctcct tgaaccctcc atggaataga 120  
 atattttaaag gagagaatgt gactcttaca tgtaatggga acaatttctt tgaagtcagt 180  
 tccaccaaat gggtccacaa tggcagcctt tcagaagaga caaattcaag tttgaatatt 240  
 gtgaatgcc aatttgaaga cagtggagaa taaaaatgtc agcaccaaca agttaatgag 300  
 agtgaacctg tgtacctgga agtcttcagt gactggctgc tcttcaggc ctctgctgag 360  
 gtggtgatgg agggccagcc cctcttcctc aggtgccatg gttggaggaa ctgggatgtg 420  
 tacaaggatga tctattataa ggatggtgaa gctctcaagt actggtatga gaaccacaac 480  
 atctccatta caaatgccac agttgaagac agtgaacct actactgtac gggcaaagt 540  
 tggcagctgg actatgagtc tgagcccctc aacattactg taataaaaagc tccgcgtgag 600  
 aagtactggc tacaattttt tatccattg ttggtggtga ttctgtttgc tgtggacaca 660  
 ggattattta tctcaactca gcagcaggtc acatttctct tgaagattaa gagaaccagg 720  
 aaaggcttca gacttctgaa cccacatcct aagccaaacc ccaaaaaaaa ctga 774

<210> 3

<211> 257

<212> PRT

<213> Homo sapiens

<400> 3

Met Ala Pro Ala Met Glu Ser Pro Thr Leu Leu Cys Val Ala Leu Leu  
 1 5 10 15

Phe Phe Ala Pro Asp Gly Val Leu Ala Val Pro Gln Lys Pro Lys Val  
 20 25 30

Ser Leu Asn Pro Pro Trp Asn Arg Ile Phe Lys Gly Glu Asn Val Thr  
 35 40 45

Leu Thr Cys Asn Gly Asn Asn Phe Phe Glu Val Ser Ser Thr Lys Trp  
 50 55 60

Phe His Asn Gly Ser Leu Ser Glu Glu Thr Asn Ser Ser Leu Asn Ile  
 65 70 75 80  
 Val Asn Ala Lys Phe Glu Asp Ser Gly Glu Tyr Lys Cys Gln His Gln  
 85 90 95  
 Gln Val Asn Glu Ser Glu Pro Val Tyr Leu Glu Val Phe Ser Asp Trp  
 100 105 110  
 Leu Leu Leu Gln Ala Ser Ala Glu Val Val Met Glu Gly Gln Pro Leu  
 115 120 125  
 Phe Leu Arg Cys His Gly Trp Arg Asn Trp Asp Val Tyr Lys Val Ile  
 130 135 140  
 Tyr Tyr Lys Asp Gly Glu Ala Leu Lys Tyr Trp Tyr Glu Asn His Asn  
 145 150 155 160  
 Ile Ser Ile Thr Asn Ala Thr Val Glu Asp Ser Gly Thr Tyr Tyr Cys  
 165 170 175  
 Thr Gly Lys Val Trp Gln Leu Asp Tyr Glu Ser Glu Pro Leu Asn Ile  
 180 185 190  
 Thr Val Ile Lys Ala Pro Arg Glu Lys Tyr Trp Leu Gln Phe Phe Ile  
 195 200 205  
 Pro Leu Leu Val Val Ile Leu Phe Ala Val Asp Thr Gly Leu Phe Ile  
 210 215 220  
 Ser Thr Gln Gln Gln Val Thr Phe Leu Leu Lys Ile Lys Arg Thr Arg  
 225 230 235 240  
 Lys Gly Phe Arg Leu Leu Asn Pro His Pro Lys Pro Asn Pro Lys Asn  
 245 250 255

Asn

<210> 4

<211> 15

<212> DNA

<213> Homo sapiens

<400> 4

tgaaatatca gattt

<210> 5  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 5  
tgaaatagca gattt

15

<210> 6  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 6  
attctgctct ccctt

15

<210> 7  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 7  
attctgccct ccctt

15

<210> 8  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 8  
gatatgatac agaaa

15

<210> 9  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 9  
gatatgacac agaaa

15

<210> 10  
<211> 15

<212> DNA  
<213> Homo sapiens

<400> 10  
tacagaaaac atttc

15

<210> 11  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 11  
tacagaatac atttc

15

<210> 12  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 12  
aattaccctt cccag

15

<210> 13  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 13  
aattaccact cccag

15

<210> 14  
<211> 15  
<212> DNA  
<213> Homo sapiens

<400> 14  
actaatgtat cctct

15

<210> 15  
<211> 15  
<212> DNA  
<213> Homo sapiens

&lt;400&gt; 15

actaatgcat cctct

15

&lt;210&gt; 16

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 16

gtatcctctc tggac

15

&lt;210&gt; 17

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 17

gtatcctatc tggac

15

&lt;210&gt; 18

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 18

taatgagcat gaatc

15

&lt;210&gt; 19

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 19

taatgagtat gaatc

15

&lt;210&gt; 20

&lt;211&gt; 15

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 20

aatcaaaaca gggtc

15